

Dissertation

Submitted to the
Combined Faculties for the Natural Sciences and for Mathematics
of the Ruperto-Carola University of Heidelberg, Germany
for the Degree of
Doctor of Natural Sciences

Presented by

Dipl.-Chem. Sabrina Annette Hanke

Born in Karlsruhe, Germany.
Oral Examination: 28.09.2015

Development of Novel Reagents for the Selective Enrichment
of Vascular Accessible Proteins and the Identification of
Disease-Specific Biomarkers

Referees:

Prof. Dr. Matthias Mayer
Dr. Christoph Rösli

The work presented in this thesis was started in September 2011 and completed in July 2015 under the supervision of Dr. Christoph Rösli in the research group Biomarker Discovery within the Division of Stem Cells and Cancer at the German Cancer Research Center (DKFZ), Heidelberg, Germany, as well as at the Heidelberg Institute for Stem Cell Technology and Experimental Medicine (HI-STEM) gGmbH, Heidelberg, Germany.

Parts of this thesis have been published in the following conference posters:

S. Hanke, A. Kerner, Y. Zhang, C. Rösli

“Hitting the Target: Novel Reagents for the Chemical-Proteomics Based Identification of Vascular Accessible Biomarkers”

ASMS 62nd Conference on Mass Spectrometry and Allied Topics 2014, Baltimore, Maryland, USA.

S. Hanke, A. Kerner, Y. Zhang, C. Rösli

“Hitting the Target: Novel Reagents for the Chemical-Proteomics Based Identification of Vascular Accessible Biomarkers”

EMBO Conference Chemical Biology 2014, Heidelberg, Germany.

S. Hanke, A. Kerner, Y. Zhang, C. Rösli

“Hitting the Target: Novel Reagents for the Chemical-Proteomics Based Identification of Vascular Accessible Biomarkers”

DGPF Proteomic Forum 2015, Berlin, Germany.

Meiner Familie.

Acknowledgements

At this point I would like to take the opportunity and thank all the people who accompanied and supported this work in various ways within the last four years.

First of all I would like to thank Dr. Christoph Rösli for his supervision and support. Thank you very much, that you were so brave four years ago to let the chemistry enter your lab. You gave me every possibility to explore the world of biology and to learn a lot of new methods and techniques. Your enthusiasm for proteomics and mass spectrometry was infectious. Thank you for leading me into this field and for letting me profit from your nearly endless knowledge. Also thank you for staying available during your Bavarian year and for proofreading my thesis. Your capability to guide through any obstacles and smaller catastrophes during every day and night time is legendary.

I would also like to thank Prof. Dr. Andreas Trumpp for giving me the opportunity to work at HI-STEM. Thank you for your trust in a project full of chemistry and mass spectrometry. I enjoyed the passionate scientific atmosphere in the lab. Also thank you for being part of my thesis advisory and my PhD examination committee.

Next, I want to thank Dr. Yixin Zhang, not only for providing us with the peptide material for the reagent synthesis, but also for your chemical guidance through this project, for your valuable advice in peptide and heparin chemistry and for proofreading parts of this thesis. Also thank you for participating in my thesis advisory committee and for travelling across Germany for those days of discussion.

I want to thank Prof. Dr. Ana García-Sáez and Dr. Martina Schnölzer for being part of my thesis advisory committee and for asking the right questions to bring forward this project. I also thank Prof. Dr. Matthias Mayer and Prof. Dr. Britta Brügger for kindly accepting the participation in my PhD examination committee.

My thank also goes to the excellent core facilities and animal caretakers of the DKFZ for providing support, knowledge and many possibilities for experiments, that would not be feasible without them.

I would like to thank the Helmholtz International Graduate School for Cancer Research for providing many opportunities to attend scientific and non-scientific seminars and courses and to meet other PhD students from different research areas.

Next, I want to thank my colleagues in the lab: Alex for his endless support in bioinformatic questions and for his bunch of tools making data analysis endurable, Wiebke for her enthusiasm for science and chocolate both enriching every discussion, Laura for bringing lots of laughter to the lab and for cheering up everyone at bad days, Amelie for making my start into the new research field easier, Renata for providing a lot of support in the first lonely ESI weeks, Katharina for her encouragement and for creating a relaxing atmosphere during lunchtime, Philipp for being the best desk partner I could imagine and for his friendly attitude bringing me down to earth if necessary, Jasmin and Kristin for our sport activities (monthly is also kind of regular) and last but not least Erika, for her never-ending support.

Zuletzt möchte ich meiner Familie danken, für jede nur mögliche Unterstützung, unendliche Geduld und Ermutigung, meinen Weg zu gehen. Mama, Papa, Oma Liane, Opa Tibor, Kathrin und Tobias, ohne euch wäre ich heute nicht hier.

Table of Contents

List of Abbreviations	17
Abstract	21
Zusammenfassung	23
1. Introduction	27
1.1. Biomarker Discovery	27
1.1.1. Biomarker – Definition, Importance and Requirements	27
1.1.2. Strategies in Biomarker Discovery	28
1.1.2.1. Disease Modelling	28
1.1.2.2. Genomics and Epigenomics	29
1.1.2.3. Transcriptomics	30
1.1.2.4. Proteomics	30
1.1.2.5. Lipidomics	31
1.1.2.6. Glycomics	32
1.1.2.7. Metabolomics and Metabonomics	32
1.2. Vascular Accessible Biomarkers	33
1.2.1. Importance of Vascular Accessible Biomarkers	33
1.2.2. Challenges in Membrane Proteomics	35
1.2.3. Identification of Vascular Accessible Protein Targets	38
1.2.3.1. Carbohydrate-Based Cell Surface Proteome Enrichment	39
1.2.3.2. Silica Coating Technology	40
1.2.3.3. <i>In Vivo</i> and <i>Ex Vivo</i> Biotinylation	41
1.3. Proteomics and Mass Spectrometry	42
1.3.1. Mass Spectrometric Approaches	42
1.3.2. Ion Sources, Mass Analyzers and Protein Identification	44
1.3.3. Quantification	47
1.3.3.1. Label-Based Quantification	47
1.3.3.2. Label-Free Quantification	48
1.3.4. Data-Independent Acquisition	50
1.4. Biomarker Validation and Targeted Therapies	52
1.4.1. Validation of Protein Biomarkers	53
1.4.2. Ligands for Vascular Targeting Applications	54
2. Aim of the Thesis	57

3. Materials and Methods	59
3.1. Chemical Synthesis	59
3.1.1. Peptide-Based Biotinylation Reagents	59
3.1.1.1. Synthesis of Biotin-(L-Asp) ₃ -β-Ala	59
3.1.1.2. Synthesis of Biotin-(L-Asp) ₃ -β-Ala-NHS and -sNHS	60
3.1.1.3. Synthesis of Biotin-(L-Asp) ₃ -Cys-CONH ₂	61
3.1.1.4. Coupling of Biotin-(L-Asp) ₃ -Cys-CONH ₂ to SMCC-Crosslinker or SM(PEG) ₆ -Crosslinker	62
3.1.2. Heparin-Based Biotinylation Reagents	63
3.1.2.1. Synthesis of Biotin-Heparin and Biotin-Enoxaparin	63
3.1.2.2. NHS- or sNHS-Activation of Biotin-Heparin and Biotin-Enoxaparin	64
3.1.3. Click Chemistry-Based Reagents	65
3.1.3.1. Synthesis of Alkyne-(L-Asp) ₃ -Cys-CONH ₂	65
3.1.3.2. Coupling of Alkyne-(L-Asp) ₃ -Cys-CONH ₂ to SMCC-Crosslinker	66
3.1.4. Commercial Biotinylation Reagents	67
3.2. Reactivity Assessment	67
3.2.1. Reaction with BSA as Model Protein	67
3.2.2. MALDI MS Linear High Mass Positive Mode	68
3.2.3. Analysis of MALDI MS Linear Mode Data	68
3.3. <i>In Vitro</i> Validation	69
3.3.1. Cell Culture	69
3.3.2. <i>In Vitro</i> Biotinylation	69
3.3.3. Visualisation of Biotinylation	70
3.3.3.1. FACS Analysis	70
3.3.3.2. Confocal Laser Scanning Microscopy	71
3.3.4. Cell Lysis	72
3.4. <i>In Vivo</i> Validation	73
3.4.1. Mouse Strain	73
3.4.2. <i>In Vivo</i> Perfusion of Mice	73
3.4.3. Visualization of Biotinylation	74
3.4.3.1. Confocal Laser Scanning Microscopy	74
3.4.4. Protein Extraction of Kidney and Liver Tissue	74
3.5. Protein Concentration Determination of Lysates and Assessment of Biotinylation	75
3.5.1. Protein Concentration Determination	75
3.5.2. ELISA	75
3.5.3. SDS-PAGE and Western Blot	76
3.6. Sample Preparation for Mass Spectrometric Analysis	77
3.6.1. Capturing of Surface Proteins out of Cell Lysates	77
3.6.2. Capturing of Vascular Accessible Proteins out of Kidney and Liver Tissue Homogenates	79

3.6.3.	Preparation of Full Proteome Samples from Kidney and Liver Tissue Homogenates	80
3.7.	Mass Spectrometric Analysis	81
3.7.1.	Discovery Experiment	81
3.7.1.1.	UPLC-Based Separation	81
3.7.1.2.	MS Acquisition (MALDI-TOF System)	82
3.7.2.	SRM Analysis	84
3.7.2.1.	Peptide Selection	84
3.7.2.2.	UPLC-Based Separation	84
3.7.2.3.	SRM Acquisition	84
3.8.	Analysis of Mass Spectrometric Data	86
3.8.1.	Data Extraction and Generation of 2D Peptide Maps	86
3.8.2.	Peptide and Protein Identification	86
3.8.3.	Determination of Peptide Proteotypicity	86
3.8.4.	Relative Quantification Using MS _Q BAT	87
3.8.5.	Propagation of Annotations Using MS _Q BAT	88
3.8.6.	Analysis of Protein Localization	88
3.8.7.	Analysis of SRM Data	89
3.8.8.	Further Data Analysis Tools	89
4.	Results	91
4.1.	Biotinylation Reagents: Chemical Entities	91
4.1.1.	Rationales and General Design of Biotinylation Reagents	91
4.1.2.	Commercial Biotinylation Reagents and Advancements	93
4.1.3.	Analysis of the Biotinylation Procedure	94
4.1.4.	Reactivity of NHS- and sNHS- β -Ala-(L-Asp) ₃ -biotin	95
4.2.	Synthesis of Novel Biotinylation Reagents	96
4.2.1.	Peptide-Based Biotinylation Reagents	96
4.2.1.1.	Design and Synthesis of Peptide-Based Biotinylation Reagents	96
4.2.1.2.	Reactivity of Peptide-Based Biotinylation Reagents	98
4.2.2.	Heparin-Based Biotinylation Reagents	99
4.2.2.1.	Design and Synthesis of Heparin-Based Biotinylation Reagents	99
4.2.2.2.	Reactivity of Heparin-Based Biotinylation Reagents	102
4.2.3.	Biotin-Free System	103
4.2.3.1.	Click Chemistry-Based Approach	103
4.2.3.2.	Synthesis and Reactivity of Click Chemistry-Based Reagents	104
4.3.	<i>In Vitro</i> Validation of Novel Biotinylation Reagents	105
4.3.1.	Workflow <i>In Vitro</i> Validation	105
4.3.2.	Visualisation of Cell Surface Biotinylation	106
4.3.2.1.	Peptide-Based Biotinylation Reagents	106
4.3.2.2.	Heparin-Based Biotinylation Reagents	107

4.3.3.	Sample Preparation for Mass Spectrometric Analysis	109
4.3.4.	<i>In Vitro</i> Validation: Mass Spectrometric Analysis	111
4.3.4.1.	Peptide-Based Biotinylation Reagents	111
4.3.4.2.	Heparin-Based Biotinylation Reagents	119
4.4.	<i>In Vivo</i> Validation of Novel Biotinylation Reagents	124
4.4.1.	Workflow <i>In Vivo</i> Validation	124
4.4.2.	Visualization of Biotinylation of Vascular Accessible Proteins .	125
4.4.2.1.	Biotinylation Depth Around Blood Vessels	125
4.4.2.2.	Biotinylation With Different Reagents	126
4.4.3.	Sample Preparation for Mass Spectrometric Analysis	128
4.4.4.	Mass Spectrometric Analysis of Biotinylated Kidney Tissue . .	131
4.4.4.1.	Discovery Experiment Kidney	131
4.4.4.2.	Background Proteome in Negative Controls	140
4.4.4.3.	SRM-Analysis of High- and Low-Abundant Proteins	143
4.4.5.	Mass Spectrometric Analysis of Biotinylated Liver Tissue . . .	147
4.4.5.1.	Discovery Experiment Liver	147
4.4.5.2.	SRM-Analysis of High- and Low-Abundant Proteins	153
5.	Discussion	157
5.1.	Design of Novel Enrichment Reagents	158
5.2.	Reactivity and Visualisation of Biotinylation	160
5.3.	Aspects of Sample Preparation for Mass Spectrometric Analysis . . .	162
5.4.	Presented Datasets and Aspects of Data Analysis	165
5.5.	The Background Proteome	168
5.6.	Biotinylation Reagents: Reactivity vs. Selectivity	171
6.	Conclusions and Outlook	175
	List of Figures	177
	List of Tables	181
	List of Schemes	183
	Bibliography	185
A.	Supplementary Figures	219
B.	Supplementary Tables	231

List of Abbreviations

7TM seven-transmembrane domain

APEX absolute protein expression

AUC area under the curve

BEMAD beta-elimination followed by Michael addition with DTT

BSA bovine serum albumin

CHCA α -cyano-4-hydroxycinnamic acid

ChIP-seq chromatin immunoprecipitation DNA sequencing

CID collision-induced dissociation

CILAT cleavable isobaric labelled affinity tag

CLSM confocal laser scanning microscopy

CSC cell surface capturing

Da Dalton

DAPI 4',6-diamidino-2-phenylindole

DBCO dibenzylcyclooctyne

DCC N,N'-dicyclohexylcarbodiimide

DDA data-dependent acquisition

DHB 2,5-dihydroxybenzoic acid

DIA data-independent acquisition

DIC differential interference contrast

DKFZ German Cancer Research Center

DMF dimethylformamide

DMSO dimethyl sulfoxide

DNA deoxyribonucleic acid

DTT dithiothreitol

ECD electron capture dissociation

ECM extracellular matrix

EDA extra domain A

EDC-HCl 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide hydrochloride

EDTA ethylenediaminetetraacetic acid

ELISA enzyme-linked immunosorbent assay

emPAI exponentially modified protein abundance index

eq equivalents

ESI electrospray ionization

ETD electron transfer dissociation
FA formic acid
FACS fluorescence-activated cell sorting
FASP filter aided sample preparation
FCS fetal calf serum
FDR false discovery rate
FFT fast fourier transform
Fmoc 9-fluorenylmethoxycarbonyl
FSC forward scatter
FTICR fourier transform ion cyclotron resonance
FWHM full width half maximum

GC gas chromatography
GEAL genetic algorithm
GO gene ontology
GPCR G protein coupled receptor
GPI glycosylphosphatidylinositol

GBTU N,N,N',N'-tetramethyl-O-(1H-benzotriazol-1-yl)uronium hexafluorophosphate
HDMS^E high-definition MS^E
HPLC high-performance liquid chromatography
HRP horseradish peroxidase

IAA iodoacetamide
iBAQ intensity based absolute quantification
ICAT isotope coded affinity tag
ICPL isotope coded protein label
IDA information dependent acquisition
IGOT isotope-coded glycosylation-site-specific tagging
IHC immunohistochemistry
IMAC immobilized metal ion affinity chromatography
IMDM Iscove's Modified Dulbecco's Medium
IMS ion mobility separation
iTRAQ isobaric tag for relative and absolute quantification

LC liquid chromatography
LC (linker) long chain
LMW low molecular weight
LOD limit of detection

m/z mass-to-charge ratio
MALDI matrix-assisted laser desorption/ionization
MES 2-(N-morpholino)ethane sulfonic acid
MOPS 3-(N-morpholino)propane sulfonic acid
MS mass spectrometry

MSX multiplexed MS/MS
MW molecular weight

NHS N-hydroxysuccinimide
NMR nuclear magnetic resonance
NP-40 4-nonylphenyl-polyethylene glycol
NSAF normalised spectral abundance factor

PA 4-pentynoic acid
PAI protein abundance index
PBS phosphate buffered saline
PCR polymerase chain reaction
PEG polyethylene glycol
PhIAT phosphoprotein isotope-coded affinity tag
PMT photomultiplier tube
PNGase peptide-N-glycosidase
ppm parts per million
PTM post-translational modification
PVDF polyvinylidene fluoride

QUIC quantitative isotopic and chemoenzymatic tagging

RIPA radioimmunoprecipitation assay
RNA ribonucleic acid
rpm revolutions per minute

S/N signal-to-noise ratio
SAGE serial analysis of gene expression
SDS sodium dodecyl sulfate
SDS-PAGE sodium dodecyl sulfate polyacrylamide gel electrophoresis
SEM standard error of the mean
SILAC stable isotope labelling with amino acids in cell culture
SISCAPA stable isotope standards and capture by anti-peptide antibodies
SLAC serial lectin affinity chromatography
SM(PEG)₆ succinimidyl-([N-maleimidopropionamido]hexaethyleneglycol)ester
SMCC succinimidyl-4-(N-maleimidomethyl)cyclohexane-1-carboxylate
sNHS N-hydroxysulfosuccinimide
SRM selected reaction monitoring
SSC side scatter
SWATH sequential window acquisition of all theoretical mass spectra

TBTA tris[(1-benzyl-1H-1,2,3-triazol-4-yl)methyl]amine
TCA tricarboxylic acid cycle
TCEP tris(2-carboxyethyl)phosphine
TFA trifluoroacetic acid
THPTA tris(3-hydroxypropyltriazolylmethyl)amine

TIPS triisopropylsilane
TIS timed ion selector
TMT tandem mass tag
TOF time of flight
TPCK tosylphenylalanylchloromethylketone
TQ tandem quadrupole
Tris 2-amino-2-hydroxymethyl-propane-1,3-diol
TUV tunable UV

UDMS^E ultra-definition MS^E
UPLC ultra performance liquid chromatography
UV ultraviolet

VEGF vascular endothelial growth factor
VIS visible

XIC extracted ion chromatogram

Abstract

A promising approach for the development of novel therapeutics with fewer side effects in healthy tissues is the targeted delivery of bioactive molecules directly to the site of disease. The prerequisite is the identification of a robust, disease-specific biomarker, targetable either with monoclonal antibodies interfering with the target's biological function or with antibody-drug conjugates delivering a therapeutic payload such as a cytotoxic drug. Hence, the target molecules of interest have to be vascular-accessible and are therefore located on the surface of diseased cells, on newly formed blood vessels or in the perivascular extracellular matrix. Proteomic approaches for the identification of novel biomarkers have to deal with (i) the high dynamic range of the proteome over at least seven orders of magnitude, (ii) the low abundance of the highly diverse plasma membrane proteome fraction, as well as with (iii) the hydrophobic character of membrane proteins. The proteins of interest are therefore often under-represented in mass spectrometric datasets of full proteome samples and cannot be stably quantified. One avenue towards the enrichment of the vascular-accessible surface proteome fraction prior to mass spectrometric analysis is the covalent modification of the target proteins with a membrane-impermeable ester-derivative of biotin, followed by streptavidin-based affinity capturing. The biotinylation of potential biomarkers is fast and efficient and can be performed by *in vitro* labelling of cells, via *in vivo* perfusion of mice or *ex vivo* using surgically resected tissue material.

The work in this thesis focused on the synthesis and multi-step validation of two novel, multiply-charged peptide-based as well as of two novel heparin-based biotinylation reagents. Furthermore, two alkyne-tagged reagents for bioorthogonal click-chemistry based enrichment were designed. The reagents' reactivity was assessed by coupling to BSA in different ratios and analysis with linear MALDI mass spectrometry. Subsequent validation was performed *in vitro* on HeLa cells and *in vivo* via perfusion of healthy NSG mice. Biotinylation efficacy was examined using FACS analysis, ELISA and Western Blot of cell and tissue samples. Cell surface or perivascular biotinylation was visualized by confocal laser scanning microscopy. Mass spectrometric analyses of the accessible proteome fractions were performed in comparison to the commercial reagents Sulfo-NHS-LC-biotin and NHS-PEG₁₂-biotin and PBS- or non-treated negative controls on biotinylated HeLa cells and kidney or liver tissue using a MALDI TOF/TOFTM 5800 system (AB SCIEX). Mass spectrometric data were analysed in terms of identified protein and proteotypic peptide numbers as well as of protein localization. Relative quantification based on MS1 signal intensities was performed using the in-house developed software MS_QBAT. SRM-analysis of some medium- and low-abundant cell surface and extracellular matrix proteins was performed in comparison to full proteome samples on a QTRAP[®] 6500 system (AB SCIEX).

The properties of any biotinylation reagent are determined by the linker between the biotin residue for affinity purification and the reactive group for protein coupling. Increase in size and polarity influences the reagents' selectivity and reactivity. Site-specific activation of the novel peptide-based biotinylation reagents could significantly improve the reactivity in comparison to non-specifically activated NHS- β -Ala-(L-Asp)₃-biotin published by Strassberger *et al.* in 2010. Alkyne-tagged reagents revealed comparable reactivity to the biotinylation reagents resulting from the similarity of the peptide-based linkers. Due to their enormous size and the increased steric hindrance, heparin-based reagents are less reactive than the smaller peptide-based and commercial reagents. The reactivity difference is also mirrored in the mass spectrometric datasets. A total of 1574 proteins could be identified within the *in vitro* analysis. The proof-of-principle study could demonstrate the stable identification of a 38-49% fraction of plasma membrane or extracellular matrix annotated proteins with the peptide-based reagents. 1965 proteins were found in the kidney dataset with a comparable fraction of 40-45% surface annotated proteins. Within the liver dataset, 1531 proteins could be identified with a slightly increased intracellular protein fraction (27-34% surface proteome fraction). Reasons for the typical background proteome identification in all samples were further assessed by different sample preparation strategies within this thesis. In short, it can be stated, that the success of biomarker studies with biotinylation reagents are dependent of the vascularisation of the target tissue to enable high biotinylation rates as well of elaborated sample preparation protocols: For example, delipidation is crucial for the work with tissue samples. A slight reactivity decrease of the novel peptide-based reagents compared to commercial Sulfo-NHS-LC-biotin was detected due to the sterically more hindered peptide linker. Nevertheless, the enrichment of the targeted surface proteome is very stable: 40% of the quantified proteins are more than 2-fold up- or down-regulated in comparable fractions between the peptide-based and the commercial reagent (kidney dataset). The number of identified proteotypic peptides per plasma membrane or extracellular matrix annotated protein is significantly increased compared to the negative controls and to the intracellular proteome fraction, which is the basis for enabling a stable protein quantification.

The mass spectrometric studies revealed that the novel peptide-based reagents provide a reliable technology platform for the enrichment of vascular accessible proteins: Plasma membrane or extracellular matrix annotated biomarker candidates can be stably identified and quantified based on high numbers of proteotypic peptides. Furthermore, stable SRM-based quantification of medium- and low-abundant targets is enabled.

Zusammenfassung

Ein vielversprechender Ansatz zur Entwicklung von neuen Medikamenten mit weniger Nebenwirkungen in gesundem Gewebe besteht im zielgerichteten Transport von bioaktiven Molekülen direkt an den Krankheitsherd. Die Entwicklung solcher innovativer Medikamente setzt zunächst die Identifizierung eines robusten und spezifischen Biomarkers als Zielstruktur im erkrankten Gewebe voraus. Diese kann im Folgenden mit monoklonalen Antikörpern, die die biologische Funktion des Biomarkers beeinträchtigen, oder mit an Antikörpern gekoppelten Medikamenten angegriffen werden. Dabei ist die Analyse des vaskulär erreichbaren Oberflächenproteoms aufgrund der leichten Zugänglichkeit durch den Blutstrom von großem Interesse: Potentielle Zielstrukturen sind auf den Zelloberflächen im erkrankten Gewebe, auf der Oberfläche neugebildeter Blutgefäße sowie in der umgebenden extrazellulären Matrix zu suchen. Die Identifizierung neuer Protein-Biomarker durch massenspektrometrische Analyse wird zum einen durch die hohe Komplexität des Proteoms mit einer Abundanz-Spanne von mindestens sieben Größenordnungen erschwert. Oberflächenproteine sind hochdivers in Struktur und Funktion, aber oftmals nur in niedriger Kopienzahl exprimiert. Des weiteren erschwert der hydrophobe Charakter der Membranproteine die Analyse. Aus diesen Gründen sind potentielle Biomarker in Vollproteom-Datensätzen meist unterrepräsentiert und können daher auch nicht stabil quantifiziert werden. Eine Möglichkeit zur Anreicherung der vaskulär erreichbaren Zielstrukturen vor der massenspektrometrischen Analyse ist deren kovalente Modifizierung mit einem Membran-impermeablen Biotin-Ester-Derivat, welches mit Hilfe von Streptavidin-Sepharose affinitätsbasiert angereichert werden kann. Die Biotinylierungsreaktion verläuft schnell und quantitativ, was neben dem Markieren von Zellen *in vitro* auch die *in vivo* Perfusion von Mäusen sowie die *ex vivo* Perfusion von humanem, operativ entferntem Gewebe ermöglicht.

Die Entwicklung und Validierung zweier mehrfach geladener Peptid- und zweier Heparin-basierter Biotinylierungsreagenzien zur selektiven Anreicherung von vaskulär erreichbaren Zelloberflächenproteinen steht im Fokus dieser Doktorarbeit. Darüberhinaus wurden zwei Alkin-modifizierte Reagenzien für bioorthogonale Anreicherung über eine Click-Reaktion synthetisiert. Die Reaktivität der Reagenzien wurde durch Reaktion mit BSA in verschiedenen molaren Verhältnissen und Analyse des modifizierten Proteins durch lineare MALDI Massenspektrometrie untersucht. Weitere Validierungsexperimente wurden mit der Zelllinie HeLa sowie mit *in vivo* perfundierten, gesunden NSG-Mäusen durchgeführt. Die Effizienz der Biotinylierung von Zell- und Gewebeproben wurde hierbei durch Durchflusszytometrie (FACS), ELISA und Western Blot Analyse bestimmt. Die Biotinylierung der Zelloberflächen bzw. die perivaskuläre Markierung mit den Reagenzien wurde mittels konfokaler Laser-Scanning Mikroskopie untersucht. Die massenspektrometrische Analyse der angereicherten Proteinfraction wurde im Vergleich zu den kommerziellen Reagenzien Sulfo-NHS-LC-biotin und NHS-PEG₁₂-biotin sowie zu PBS- oder unbehandelten

Negativkontrollen durchgeführt. Die angereicherten Proteinlysate von biotinylierten HeLa-Zellen sowie von Nieren- und Lebergewebe von perfundierten Mäusen wurden dabei auf einem MALDI TOF/TOF™ 5800 System (AB SCIEX) analysiert. Die anschließende Datenanalyse konzentrierte sich auf die vergleichende Untersuchung der Anzahl der identifizierten Proteine und proteotypischen Peptide, sowie deren subzelluläre Lokalisierung. Eine relative Quantifizierung wurde basierend auf den MS1 Signalintensitäten mit Hilfe der Labor-eigenen Software MSQBAT durchgeführt. Mittels SRM-Analyse wurden darüber hinaus einige Oberflächenproteine mit niedriger oder mittlerer Abundanz im Vergleich zu nicht angereicherten Vollproteom-Proben auf einem QTRAP® 6500 System (AB SCIEX) quantifiziert.

Die Eigenschaften von Biotinylierungsreagenzien werden durch den Linker zwischen dem Biotinrest und der reaktiven Gruppe zur Modifizierung der Zielproteine bestimmt. Eine Vergrößerung des Reagenz oder eine Veränderung der Polarität beeinflussen die Reaktivität und die Selektivität. Die regioselektive Aktivierung der neuen Peptid-basierten Reagenzien erhöht die Reaktivität im Vergleich zum unspezifisch aktivierten NHS- β -Ala-(L-Asp)₃-biotin (Strassberger *et al.*, 2010) signifikant. Die alkinmodifizierten Peptid-Reagenzien weisen eine ähnliche Reaktivität wie die strukturell verwandten Biotinylierungsreagenzien auf. Aufgrund ihrer enormen Größe und der daraus resultierenden sterischen Hinderung sind die Heparin-basierten Reagenzien weniger reaktiv als die wesentlich kleineren Peptid-basierten oder die kommerziell erhältlichen Reagenzien. Diese Reaktivitätsdifferenz spiegelt sich auch in der massenspektrometrischen Analyse wider. Im Gesamten konnten in der *in vitro* Analyse 1574 Proteine identifiziert werden. Das Pilotexperiment konnte belegen, dass die Peptid-basierten Biotinylierungsreagenzien die stabile Identifizierung einer 38-49% großen Fraktion von Proteinen ermöglicht, die in der Plasmamembran oder der extrazellulären Matrix lokalisiert sind. Im Nieren-Datensatz wurden 1965 Proteine mit einer vergleichbaren Fraktion an Oberflächenproteinen von 40-45% gefunden. Die Identifizierung intrazellulärer Proteine war im Leber-Datensatz leicht erhöht: 1531 Proteine mit einem Oberflächenproteomanteil von 27-34%. Die für solche Datensätze typische Identifizierung eines intrazellulären Hintergrundproteoms wurde in dieser Arbeit mittels verschiedener Probenaufarbeitungs-Strategien ebenfalls untersucht. Zusammenfassend lässt sich feststellen, dass der Erfolg von Biomarker-Identifizierungsstudien zum einen von der Vaskularisierung des Zielgewebes abhängig ist, um hohe Biotinylierungsraten zu ermöglichen, und zum anderen von verfeinerten Protokollen zur Probenpräparation: Beispielsweise erwies sich die Delipidierung als wesentlich für die Arbeit mit Gewebeproben. Trotz der verminderten Reaktivität der Peptid-basierten Reagenzien gegenüber des kommerziellen Sulfo-NHS-LC-biotin durch den sterisch anspruchsvolleren Peptidlinker verläuft die Anreicherung des Zielproteoms sehr stabil: 40% der quantifizierten Proteine sind mehr als 2-fach reguliert, und zwar zu gleichen Teilen zwischen Peptid- und kommerziellem Reagenz (Nieren Datensatz). Die Anzahl der identifizierten proteotypischen Peptide von Plasmamembran oder extrazellulärer Matrix-annotierten Proteinen ist im Vergleich zu den Negativkontrollen und der intrazellulären Proteinfraction signifikant erhöht - die Basis für jede erfolgreiche Quantifizierung.

Die massenspektrometrischen Validierungsstudien konnten zeigen, dass die neuen Peptid-basierten Biotinylierungsreagenzien eine verlässliche Technologie zur Anreicherung vaskulär erreichbarer Proteine darstellen: Plasmamembran- oder extrazelluläre Matrix-Proteine konnten basierend auf einer Vielzahl von proteotypischen Peptiden stabil identifiziert und quantifiziert werden. Darüber hinaus ermöglichen die neuen Reagenzien auch eine robuste SRM-basierte Quantifizierung von Zielproteinen mit mittlerer bis niedriger Abundanz.

1. Introduction

In this chapter, an introduction to the field of biomarker discovery is given, in which this thesis is embedded. The introduction covers the entire range from basic principles in biomarker discovery to the application of biomarkers in targeted therapies. Hereby, the focus is on vascular accessible biomarkers and mass spectrometry based proteomics, reflecting the focus of this thesis.

1.1. Biomarker Discovery

1.1.1. Biomarker – Definition, Importance and Requirements

According to the Biomarker Definitions Working Group from the National Institutes of Health, USA (NIH) a biomarker or biological marker is “a characteristic that is objectively measured and evaluated as an indicator of normal biological processes, pathogenic processes, or pharmacologic responses to a therapeutic intervention”.^[1] Biomarkers serve as diagnostic tool for the detection of a disease, to evaluate the stage of the disease and disease survival (prognostic biomarker), to classify subtypes or to identify patients, that benefit from a specific therapy (predictive biomarker), and to monitor disease progression or a therapeutic response.^{[1],[2]} Hereby, according to the World Health Organisation (WHO), “a biomarker is any substance, structure or process that can be measured in the body or its products and influence or predict the incidence of outcome or disease”.^[3] Examples span from simple measurement of blood pressure or imaging of tumour size or blood vessel growth, to molecular biomarkers such as the glucose level in blood or antigens specifically expressed in a diseased tissue.^{[1],[4],[5]}

Molecular differences between healthy and diseased tissue or across different subgroups of a disease are assessed to gain mechanistic insights in disease development and progression or to exploit the detected changes for diagnostic or therapeutic purposes. Changes include alterations in expression or molecular modifications, hereby DNA, RNA, proteins, lipids, carbohydrates or a metabolites can serve as biomarkers.^{[6],[7],[8]} A promising biomarker candidate needs to meet specific sensitivity and specificity criteria: Sensitivity quantifies the probability for correct detection of the disease (true positive rate), specificity measures the likelihood for correct identification of negatives (true negative rate).^[9] For clinical relevance, a 90%/90% standard is recommended.^[10]

Besides the prognostic, predictive and monitoring power of a biomarker, molecules specifically expressed on the diseased tissue can also be exploited for targeted thera-

pies.^[2] According to the National Institutes of Health, a targeted therapy is “a type of treatment that uses drugs or other substances to identify and attack specific types of [...] cells with less harm to normal cells”.^[11] For vascular targeting of a diseased tissue with a diagnostic or a drug, both a disease biomarker and a binding molecule to that biomarker, e.g. an antibody, have to be identified, and, depending on the mode of action, a bioactive moiety has to be coupled to the binder.^[12] Requirements for a high-quality biomarker include up-regulation in the diseased tissue, and ideally absence in healthy tissue, a stable and reasonably abundant expression as well as accessibility for the drug.^{[12],[13]} Also differentially regulated splice isoforms can serve as biomarkers.^{[14],[15]} Furthermore, protein translocation can be detected under specific disease conditions, which might not result in deregulation of protein abundance, but in differential accessibility for drugs.^{[16],[17],[18],[19],[20]}

1.1.2. Strategies in Biomarker Discovery

1.1.2.1. Disease Modelling

Model systems to recapitulate human disease are of crucial importance in biomarker research. Human material is not accessible in the required quantities.^[9] Furthermore samples can be highly heterogeneous among different populations and individuals, e.g. in terms of the molecules’ abundances, structural variants or modifications.^[21] Despite the genetic background, also other factors such as environment, lifestyle, age or sex affect a clinical sample.^{[8],[9]} Further diversity in diseased cohorts might occur due to different stages of disease, treatments obtained or molecular subtypes.^[9] For those reasons, most biomarker investigations begin with a model system.

Cell lines represent a valuable *in vitro* model: As the culture system can be well-defined, differences and changes upon manipulation can be detected reliably. Many commonly used cell lines are immortalised to obtain unlimited proliferation capacity. Immortalisation is associated with tumourigenesis, therefore the cell phenotype might slightly change. Long-term passaging of the cell lines additionally leads to genetic alterations and clonal selection.^[22] As it is questionable whether the phenotype of immortalized cells still reliably recapitulates the disease situation, primary cell lines directly derived from resected patient material might be the better *in vitro* model. However, successful culture conditions for those cells are more difficult to establish.^[23] To examine diseases in an *in vivo* situation, genetically engineered mouse models are commonly used. Hereby, the gene of interest might either undergo knockdown or knockout (loss of function), or being overexpressed or introduced (so-called transgenesis) (gain of function).^[24] Whereas genetically engineered animal models are very useful to examine the effect of a mutation in a specific gene, xenograft models are favoured to mimic human disease. Human tissue material (e.g. primary tumour material) is transferred into immunocompromised animals such as NOD/SCID mice.^[25] Within xenograft models, the heterogeneity of the patient material as well as the interaction with the microenvironment are more realistically reproduced or mimicked. Furthermore, many molecular features of the diseased

tissue are conserved.^[26] Finally, *ex vivo* biomarker studies of patient material are performed, either in surrogates such as blood (secreted biomarkers) or with surgically resected tissue material.^[27] In the following, biomarker discovery approaches on different molecular levels (see Figure 1.1) are further elucidated.

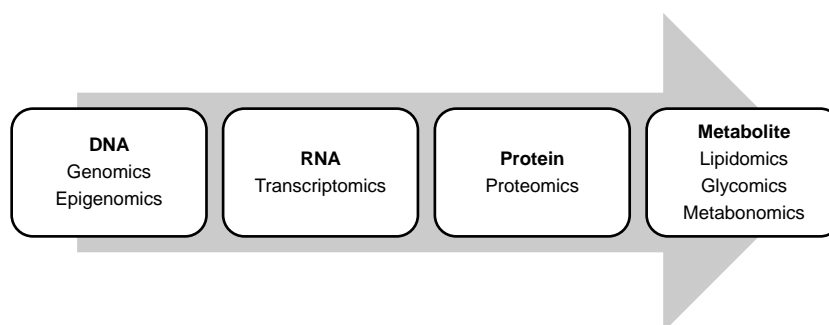


Figure 1.1.: Biomarker discovery on different molecular levels.

1.1.2.2. Genomics and Epigenomics

Many human diseases are characterized by alterations of the genome including mutation of the DNA, copy number variations, chromosomal aberrations or epigenetic changes.^{[28],[29]} Since Sanger developed the first DNA sequencing method in the 1970s, considerable efforts have been undertaken to determine the whole DNA sequence of organisms.^{[30],[31]} Next-generation sequencing technologies perform the analysis of genomes in a heavily parallelized way and are about 200 times faster and cheaper than the conventional Sanger method. In the course of next-generation sequencing, single-stranded DNA molecules are immobilised on a solid support such as a glass slide or beads and the sequencing is performed with modified pyrosequencing protocols.^{[32],[33]} Sequencing machines are meanwhile commercially available from many companies.^[32]

Epigenetic regulations such as DNA methylation or histone modification do not affect the DNA sequence but influence the chromatin structure and gene transcription.^[29] For example, CpG island methylation in promotor regions occurs in normal cells to down-regulate gene expression by prevention of transcription factor or RNA polymerase binding. In early tumour development, the DNA methylation pattern often changes, e.g. resulting in the prevention of tumour suppressor gene expression by hypermethylation of the promotor regions.^{[29],[34]} Common chemical histone modifications are methylation/demethylation or acetylation/deacetylation, resulting in changes of the chromatin structure and differential gene regulation.^[34] Analysis techniques include bisulfite sequencing to examine the methylation status of cytosines or chromatin immunoprecipitation DNA sequencing (ChIP-seq) to study histone modification and nucleosome positioning.^[35]

1.1.2.3. Transcriptomics

Transcriptomics describes the analysis of the RNA molecules transcribed from the genome, including mRNA, small RNAs and non-coding RNAs, whose expression indirectly influences most processes within a cell in health and disease.^[36] Analysis techniques are either relying on hybridisation or on sequencing.^[36] Microarray-based gene expression profiling is conducted by hybridisation between single-stranded DNA probes of defined sequences immobilised on a solid support and differentially fluorescently labelled cDNA derived from the samples' RNA by reverse transcription.^[37] Hybridisation techniques are high-throughput and relatively cheap, but limited by the prior selection of the detectable sequences and experimental challenges such as non-specific cross-hybridisation or the accessible dynamic range.^[36] The serial analysis of gene expression (SAGE) technology, developed for the large-scale quantification of transcripts, is a tag-based sequencing approach. In brief, short nucleotide sequence tags, able to clearly identify the corresponding transcripts, are generated and ligated to so-called concatemers, amplified by PCR and sequenced.^{[38],[39]} The development of high-throughput DNA sequencing methods provided the platform for the development of high-throughput RNA-Seq for whole transcriptome profiling. Hereby, a library of cDNA fragments is generated and subsequently sequenced.^[36]

Whereas some circulating miRNAs either derived from miRNA containing microvesicles or from cell debris were found to be potent blood-based biomarkers for several cancer types, it is questionable if the whole transcriptomic profile can directly mirror the biological phenotype.^{[40],[41]} Whereas RNA-Seq can detect the expressed genes, splicing isoforms and novel or rare transcripts, the data cannot provide information about translational regulation, protein degradation or posttranslational modifications, e.g. phosphorylation or glycosylation.^[41] The median half-life of proteins is reported to be around five times as long as for mRNA (46 h vs. 9 h, respectively).^[42] The dynamic range of the transcriptome covers about three to four orders of magnitude, whereas the proteome of a cell spans a range of about seven orders of magnitude.^[42] Translational regulation of protein expression was found as main cause for differences in protein abundance.^[42] A large-scale study of the human proteome supports this fact by showing that the ratio of mRNA and protein level for one specific protein is well-conserved across different tissue types, whereas absolute abundance levels and ratios might heavily vary.^[43]

1.1.2.4. Proteomics

The analysis of the proteome, which includes protein abundance, structure, modification, interaction and function, is crucial for the understanding of health and disease and the development of novel diagnostics or therapeutics.^[44] In general, two methodologies are applied to study the proteome: Affinity- or antibody-based techniques and mass spectrometric approaches.^[45] Despite the clinically applied antibody-based methods immunohistochemistry (IHC) of paraffin embedded tissue samples or enzyme-linked immunosorbent assay (ELISA) using body fluids or tissue lysates to de-

test or quantify single proteins, many attempts have been undertaken to multiplex this type of analysis. Tissue microarrays are commonly applied to screen hundreds of tissue specimens simultaneously.^[46] Cytometric bead arrays are based on different antibody coated beads with diverse fluorescent properties mixed with cell or tissue lysates and analysed all at once via FACS. Microtitre plate-based arrays or slide-based systems with robotically deposited antibodies (more than 2000 features per cm²) exploit and multiplex the classical sandwich ELISA principle with fluorescent detection.^[47] Antibody-based technologies suffer from the cross-reactivity of antibodies and the limited sensitivity and specificity of immunoassays.^[45] Importantly, even the standards in commercial antibody kits are not necessarily reproducible.^[47] Therefore, reliable comparisons are difficult to perform.^[47] Mass spectrometric approaches are currently the method of choice for large-scale proteomic studies. As the focus of this thesis is on mass spectrometry-based proteomics, methods and technologies are further described in a distinct section (see chapter 1.3).

1.1.2.5. Lipidomics

Since the mid-2000s, the molecular profiling of whole lipidomes from organisms, cells or cell organelles and the deciphering of cellular lipid networks developed into the rapidly growing research field of lipidomics.^{[48],[49]} Lipid formation is in most cases based on condensation of esters or isoprene units.^[49] The high structural diversity of lipids is generated by combination of fatty acids, headgroups and backbones, e.g. glycerol.^[50] The character of lipids hereby reaches from high hydrophobicity to amphiphilicity. The main lipid classes include fatty acyls, glycerolipids, glycerophospholipids, polyketides, sterol lipids, prenol lipids, sphingolipids and saccharolipids.^{[49],[50]} Lipid function does not only consist in dynamic barrier formation in form of membranes. Lipids are also involved in energy homeostasis and signalling.^{[50],[51]} Changes in the lipid profiles have been reported for different nutrition conditions, e.g. dietary fat consumption, as well as for many diseases, such as cancer, diabetes or neurodegenerative diseases.^{[52],[53],[54],[55],[56]} For analysis, the hydrophobic lipids have to be firstly extracted from the source (cells/tissue), most often by phase separation, e.g. via lipid extraction in organic solvents, a method that can be also adapted to isolate highly hydrophobic lipid modified peptides.^{[57],[50],[58]} Chromatographic separation methods, such as high-performance liquid chromatography (HPLC), gas chromatography (GC) or thin-layer chromatography are well established for all different kinds of lipids.^[50] Analyses in lipidomics usually rely on mass spectrometry, mostly via electrospray ionisation (ESI).^{[57],[59]} Furthermore, mass spectrometric imaging of lipids on intact tissue specimens is reported via matrix-assisted laser desorption/ionization (MALDI).^[60] Non-destructive ¹H- or ³¹P-NMR measurements can be used for structural analyses of single compounds.^[50]

1.1.2.6. Glycomics

Glycomics is the study of the glycan profile of an organism. Glycans are found as dense coating of the cell surface in the form of glycoproteins or glycolipids.^[61] About half of the proteome of a cell is estimated to be glycosylated.^[62] The glycan signature of a cell plays an important role in the immune system, e.g. for pathogen recognition, in cell-cell contacts, in the mediation of cell adhesion and migration or in signalling.^{[61],[63]} Modified glycosylation patterns are found under pathogenic conditions. For example, in the majority of cancers especially fucosylation and sialylation are reported to be deregulated.^{[64],[65]} The high diversity of the sugar residues arises from the enormous combinatorial possibilities: The human glycome is mainly constructed out of nine monosaccharide building blocks, which can be differentially linked, both at different positions and under different stereochemistry, to form linear or branched structures.^[61] Attachment to proteins occurs via N- or O-glycosidic linkages to different amino acid side chains.^[61] As about 75% of protein linkages are reported to be exclusively N-glycosidic, and further 10% of the proteins contain both type of linkages, N-glycan release with PNGaseF is a commonly used step in sample preparation.^{[62],[64]} For structural analysis via mass spectrometry, different enzymatic depolymerization methods for the glycosyl chains are reported.^[66] In general, analysis methods include chromatographic separation, ESI- or MALDI-based mass spectrometry as well as structural analysis (monosaccharide composition and linkage information) via NMR.^[63] Furthermore, carbohydrate microarrays have been developed to screen for binding specificity to proteins.^[67] Methods for glycoproteome enrichment and analysis are further described in chapter 1.2.3.1.

1.1.2.7. Metabolomics and Metabonomics

Metabolomics describes the measurement of the concentrations of intermediates and products of the metabolism. Metabonomics can be understood as a subset of metabolomics and deals with quantitative profiling of metabolite changes upon pathological stimuli, genetic effects, changes in diet, environment or pharmaceutical intervention.^{[68],[69]} The assessed metabolites are mostly small molecules ($M < 1000$ g/mol) such as amino acids, vitamins, nucleotides or organic acids.^{[69],[70]} The studied material reaches from a wide range of biofluids (e.g. urine, blood, cerebrospinal fluid) to extracts from tissue biopsies.^[68] One common analysis technique is nuclear magnetic resonance (NMR), a non-destructive and rapid measurement used for metabolite fingerprinting.^[70] Further analysis methods are based on mass spectrometry, either coupled to liquid or gas chromatography or to capillary electrophoresis.^{[70],[71]}

1.2. Vascular Accessible Biomarkers

1.2.1. Importance of Vascular Accessible Biomarkers

A monoclonal antibody or an antibody coupled to a diagnostic or a therapeutic payload cannot access all possible biomarkers identified by comparative analysis of a whole tissue samples. The diffusion characteristics of a vascular administered antibody are mainly determined by its size and charge characteristics and the affinity for the target. Small antibodies with low affinity and a low target number in the surrounding tissue are able to reach a larger penetration depth around blood vessels, however for the price of reduced retention in circulation and faster clearance.^{[72],[73]} Additionally, in tumour tissue, the elevated interstitial pressure as well as proliferating cancer cells compressing the tumour vessels further reduce the extravasation rates of monoclonal antibodies.^{[74],[75],[76]} Moreover, the vascular density in tumours expressed as capillary surface area-to-tumour volume ratio is often reduced and the blood flow irregular leading to a heterogeneous drug distribution within the tumour.^{[77],[78],[74],[79]} Areas more than 100 μm distant from blood vessels are therefore normally not accessible for monoclonal antibodies.^[78]

The vascular endothelium is not just a physical barrier around blood vessels: Functions under physiological and pathological conditions include the active transport of small molecules, regulation of blood pressure and coagulation, adhesion or transmigration of inflammatory cells and the formation of new blood vessels.^{[80],[81]} Angiogenesis describes the process of blood vessel sprouting out of pre-existing vessels and is a rare event in an adult, occurring in wound healing, during the menstrual cycle, and in pathological situations.^[82] Examples for angiogenesis under pathological conditions are further elucidated in the following.

During solid tumour development, increasing nutrient deprivation and hypoxia activate the so-called angiogenic switch, as supply with oxygen and metabolite transport is essential for tumour growth.^{[83],[84]} Physical inhibition of tumour nodule vascularisation in rodents led to a restriction of the tumour diameter to approximately 0.4 mm.^[83] Microscopic tumours with less than 1 mm³ volume are normally avascular and dormant, and are in most cases neither diagnosed nor become harmful for the host.^{[85],[86]} The point in time of the angiogenic switch depends on the tumour type and its microenvironment, factors like vascular endothelial growth factor (VEGF) stimulate the sprouting of blood vessels.^{[84],[87]} Tumour vasculature has an abnormal architecture compared to the normal vasculature. The aberrant vascular structure is characterised by a hazard and chaotic pattern of interconnections, vessels are leaky, the endothelial cell monolayer is defective and providing gaps, pericytes stabilizing the capillary structure are only loosely attached or absent and the basement membrane is either absent or too thick.^{[87],[79],[88]} Tumour endothelial cells are irregular in morphology, e.g. exhibiting cytoplasm-filled projections, and are also distinct at the molecular level as shown by gene expression and proteomic analyses.^{[88],[89],[90],[91]}

Cell lines are often used to model disease *in vitro*, but it is well-accepted that

there is a gap between the model and the *in vivo* situation.^[92] In a comparative proteomic study using the enriched endothelial fraction of rat lung tissue vs. *in vitro* cultured rat lung microvascular endothelial cells, 41% of the *in vivo* expressed plasma membrane proteins could not be detected in the *in vitro* situation.^[93]

Besides the prominent example of tumour formation, there are also other pathological situations with angiogenic events.^[94] Chronic inflammatory diseases such as inflammatory bowel or inflammatory liver disease are also reported to promote angiogenesis. Inflammatory tissue is often hypoxic, favoring vessel ingrowth and the high mass of inflammatory cells (e.g. macrophages, lymphocytes, mast cells) further stimulates angiogenesis.^{[95],[96],[97]} Furthermore, the high blood flow during inflammation induces shear stress on the endothelium. The increased endothelial surface can further promote angiogenesis to ensure supply with nutrients and further inflammatory cells.^{[95],[96]} Also rheumatoid arthritis is linked to angiogenesis. The progressive disease is characterised by hyperplasia of the synovial membrane and the influx of leukocytes and inflammatory cells.^[98] The formation of a pannus, i.e. an abnormal fibrovascular tissue layer, is highly dependent on vascularisation.^{[99],[100]} Furthermore, the pathogenesis of psoriasis is driven by angiogenesis. The skin is massively infiltrated with immune cells, vessels enlarge, and dermal vascularity increases.^{[94],[101]} Uncontrolled ocular angiogenesis occurs in retinopathy of prematurity (ROP) in children or in proliferative diabetic retinopathy (DR). Fragile blood vessels are formed leading to leakage of blood (hemorrhage) and fluid accumulation in the ocular cavity. If untreated, consequences include intransparency of the cornea, formation of macula edema, irreversible damage of the retina and impairment of retinal neurons, finally resulting in blindness.^{[102],[103]} In cardiovascular medicine, many preclinical and clinical studies evaluate the potential of angiogenic therapies, e.g. the delivery of angiogenic growth factors such as VEGF to the ischaemic tissue to stimulate the growth of novel blood vessels.^{[94],[104],[105]}

As angiogenesis is a characteristic of many pathogenic situations, the identification of targets on the angiogenic vasculature is of greatest interest for the development of novel therapies. Monoclonal antibodies targeting vascular biomarkers can on the one hand interfere with angiogenic processes and disrupt blood flow and supply with nutrients or oxygen in the diseased tissue mass.^{[91],[106]} On the other hand, an approach appearing contradictory at first sight was developed. Vascular normalisation, i.e. recovery of normal blood vessel architecture and function, can be achieved with certain antiangiogenic agents for example targeting the prolylhydroxylase domain 2 (PHD2) protein. This strategy is used to improve drug delivery to the diseased tissue mass.^{[107],[79],[108]} Furthermore, the application of antibody conjugates with diagnostic or therapeutic payloads is possible. Prostate-specific membrane antigen (PSMA) is a membrane-bound glycoprotein expressed in benign prostate tissue and with higher abundance rates also in prostate cancer. The antigen is also found at high levels on tumour neovasculature, not only in prostate cancer, but also in clear cell renal carcinoma, colonic adenocarcinoma and non-small cell lung carcinoma, whereas no expression was detectable in normal vasculature.^[109] A monoclonal anti-PSMA antibody coupled to Indium 111 has recently been evaluated in a phase-I

trial for targeted imaging of progressive solid tumours.^[110]

The extracellular matrix (ECM) is the non-cellular component of all tissues, providing not only a physical scaffold, but also an influential microenvironment for cells. The main components of the extracellular matrix are fibrous proteins (e.g. fibrillar collagens or fibronectin) as well as proteoglycans. Importantly, the exact biochemical composition is tissue-specific and highly heterogeneous.^[111] Dynamic remodelling of the extracellular matrix is a regulatory mechanism interfering with many processes such as cell differentiation and angiogenesis. Remodelling can include compound degradation, rearrangement or modification e.g. by matrix metalloproteinases (MMPs).^[112] Anomalies occur in many diseases: Deregulation in cancer contributes to disease progression, furthermore the extracellular matrix plays a role as niche for cancer stem cells.^{[113],[114],[112]} Angiogenesis promotes local extracellular matrix degradation to enable the migration of endothelial cells into the perivascular space.^[87] Gene expression profiling of colorectal tumour-associated endothelium revealed several deregulated extracellular matrix proteins.^[89] In hypertensive heart disease, cardiac fibrosis occurs due to MMP inhibitor deregulation.^[115] Extracellular matrix changes are also involved in fibrosis development. The excessive accumulation of fibrous connective tissue is observed in many diseases such as chronic obstructive pulmonary diseases or hepatic diseases.^[116] Also in inflammatory diseases remodelling occurs. The alternatively spliced EDA domain of fibronectin has been identified as promising biomarker for tissue remodelling and angiogenesis in inflammatory bowel disease. Anti-inflammatory activity of an Interleukin-12-coupled antibody against the EDA domain could recently be shown in a colitis mouse model.^[14] The corresponding Interleukin-10 and Interleukin-4 derivatives have also been shown to reduce arthritic progression, chronic skin inflammation and endometriotic lesion development.^{[15],[117],[118]}

Finally, to put it all in a nutshell, for the identification of novel, vascular accessible biomarkers, proteins expressed on the surface of diseased cells, in the surrounding extracellular matrix or on newly formed blood vessels have to be analysed.

1.2.2. Challenges in Membrane Proteomics

Membranes do not only play a role as physical barrier in cell architecture. Proteins associated with the lipid bilayers of the plasma membrane or the organelle membranes (e.g. Golgi apparatus, endoplasmic reticulum, mitochondria, nucleus) play important roles in cellular communication, signal transduction and the transport of small molecules, metabolites and ions.^[119] The membrane proteome is hereby defined as the “entire complement of membrane proteins present in a cell under a specific condition and at a specific time”.^[119]

Membrane proteins have an amphipathic character, hydrophobic protein regions hereby interact with the hydrophobic lipid tails. Membrane spanning domains contain sequences of 15-25 hydrophobic amino acids and mainly form transmembrane α -helices or β -barrel structures.^[120] The following classes of membrane proteins can be distinguished (see Figure 1.2):^{[120],[121],[122]}

- Type I single-pass transmembrane protein: C-terminus at cytoplasmic side.
- Type II single-pass transmembrane protein: N-terminus at cytoplasmic side.
- Multipass transmembrane protein: typically with 2 to 20 membrane-spanning domains, prominent example: seven-transmembrane domain (7TM) receptors.
- Peripheral membrane proteins: attached to the membrane surface or to integral membrane proteins.
- Lipid chain anchored membrane proteins: proteins covalently bound to a lipid anchor.
- GPI anchored membrane proteins: proteins covalently bound to a glycosylphosphatidylinositol anchor.

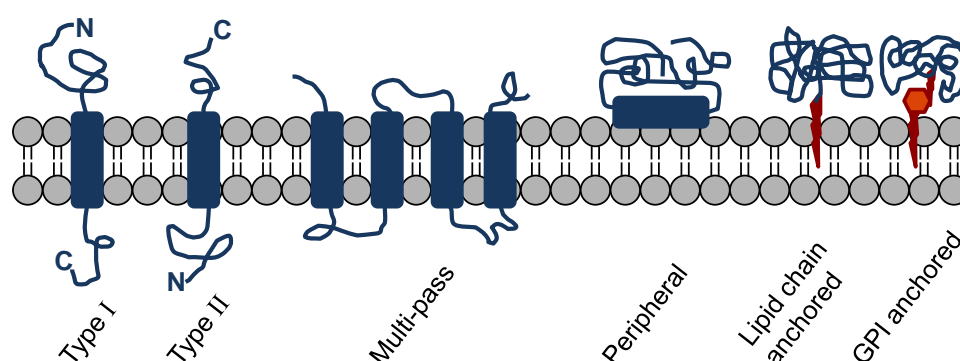


Figure 1.2.: Schematic illustration of different classes of membrane proteins.

Recent estimations report 27% of the human proteome to exhibit α -helical structural elements, generally most studies map about 30% of the human genome to membrane proteins.^{[123],[124],[125]} Based on their function, membrane proteins can be classified into three major groups: Receptors, transporters and enzymes.^[123] Membrane proteins account for about 70% of all known drug targets.^{[119],[126]} Due to their accessibility for drugs and the functional involvement in many signal cascades, the plasma membrane localised G protein coupled receptors (GPCR or 7TM receptors) are one of the most potent drug target families on the market.^{[122],[127],[128],[129]} Approximately 25% of all currently known drugs affect GPCRs.^[127]

The proteome of a mammalian cell exhibits a high dynamic abundance range of at least seven orders of magnitude.^[41] Typically about 10 000-12 000 genes are expressed within a cell.^[41] The median copy number of a protein per cell is estimated to about 18 000, the forty most abundant proteins (e.g. Histone proteins, Filamin A, Pyruvate kinase, Heat shock protein 60) contribute about 25% of the whole protein mass. The 600 most abundant proteins account for 75% of the total protein mass, whereas the lower half of the proteome only constitutes to less than 2%.^[130] Ribosomal components and the core proteasome subunits as well as many structural

proteins of the cytoskeleton belong to the most abundant proteins within a cell.^[130] Metabolic enzymes occur across the whole abundance range: Glycolytic enzymes can account for up to 10% of the protein mass, TCA cycle enzyme expression is about one order of magnitude below. Components of amino acid metabolism or of fatty acid synthesis are far lower expressed.^[131] In a study performed on HeLa cells, about 25% of all proteins were found to be integral to membranes, with only a 7.6% fraction in terms of the total proteome mass.^[130] About 10% of the proteins were annotated to the plasma membrane.^[130]

The good accessibility and the wide range of functions make the cell surface proteins prominent biomarker candidates. The analysis of the mostly low-abundant plasma membrane proteins is challenging. As indicated above, proteomic approaches have to deal with the enormous sample complexity and the high dynamic range of the proteome, leading to an under-representation of the target plasma membrane proteome fraction in full proteome datasets.^{[120],[132],[119],[133]}

The hydrophobic character of plasma membrane proteins is another challenge in membrane proteomics: Efficient solubilisation and the prevention of aggregation or precipitation is crucial for the analysis.^{[120],[132],[119]} Solubilisation approaches include the application of detergents such as SDS or Triton X-100 to disrupt membrane protein interactions as well as treatment with organic solvents, e.g. by sonication in 60% methanol or digest in 90% formic acid.^{[134],[135],[136],[137]} As most detergents negatively interfere with mass spectrometric analysis by decreasing the chromatographic resolution and suppressing ionisation, a detergent removal step has to be included in every sample preparation strategy.^{[138],[135]} In full proteome samples, solubilisation agents can be removed by protein precipitation, detergent removal resin columns or with the filter-aided sample preparation (FASP) protocol using urea to disrupt detergent micelles.^[139] Acid-labile anionic surfactants such as sodium 3-[(2-methyl-2-undecyl-1,3-dioxolan-4-yl)methoxyl]-1-propanesulfonate commercialised under the name RapiGestTM (waters) were developed to aid protein solubilisation without interference with the LC/MS analysis. At low pH, the denaturant is cleaved into non-interfering side-products. The water-immiscible tridecan-2-on can be removed by a centrifugation step.^[140] Although an improved solubilisation and mass spectrometric identification of membrane proteins is reported, the technique might suffer from co-precipitation of hydrophobic proteins with the hydrophobic surfactant by-product.^{[140],[139]}

As the hydrophobic transmembrane domains lack polar amino acid side chains such as arginine and lysine, the digest of those segments with the commonly used enzymes trypsin or Lys-C does not result in short peptide fragments required for a proper mass spectrometric analysis.^{[120],[132],[119]} As the methionine content is increased in transmembrane regions, chemical cleavage with cyanogen bromide (cleavage C-terminal of methionine) subsequently to tryptic digest was suggested.^[141] In a study on a single transmembrane protein, cyanogen bromide could be shown to double the cleavage in transmembrane domains, whereas the cleavage of the hydrophilic domains was not significantly changed.^[142]

Several protocols to analyse membrane proteins are based on different shaving

methods.^[133] Peripheral membrane proteins can be removed from the cell surface by washing in buffers of high ion strength or of changing pH (also termed carbonate stripping).^{[143],[139]} To remove plasma membrane proteins anchored by a GPI moiety, phospholipase D treatment was applied.^[144] Proteinase K treatment of cells could be shown to shave the extracellular hydrophilic protein regions for analysis.^[145] Basis for all those methods is the preservation of the cell membrane integrity during the procedure.^[133]

Density gradient centrifugation has been widely exploited for the step-wise enrichment of different cell organelles and also for the plasma membrane fraction.^{[120],[146]} Nevertheless, cross-contaminations between different organelle membranes (e.g. plasma membrane, mitochondrial membrane, endoplasmic reticulum membrane) as well as non-covalent attachment of highly abundant intracellular proteins to membrane sheets occurs as undesired side-effect.^[133] Using an optimized protocol for plasma membrane proteome preparation relying on sequential density-gradient based centrifugations steps, about half of the identified proteome fraction was membrane annotated, but only about 200 plasma membrane proteins could be identified (20% fraction).^[147]

Further enrichment technologies for the surface proteome fraction will be elucidated in the following chapter, in combination with the biomarker criterion of vascular accessibility.

1.2.3. Identification of Vascular Accessible Protein Targets

Early vascular accessible, disease-specific biomarkers owe their discovery extensive immunohistochemical analyses, e.g. with antibodies derived from rodents immunized with purified endothelial cells from tumours, or sometimes just fortunate coincidences.^{[148],[149],[150],[151]}

Inspired from the phage display technology, *in vivo* selection of phage libraries was developed in the late 1990s to select and isolate peptides specifically binding to vascular targets in the diseased tissue.^{[152],[153]} Peptide phage libraries contain billions of different bacteriophages displaying a polypeptide sequence on their surface directly linked to their engineered DNA sequence.^{[154],[155]} The peptide phage libraries are injected into rodent disease models and, after resection of the tissue of interest, vascular binders are isolated, amplified and identified or subjected to the next round of panning.^{[152],[155]} The technique has been applied to identify peptides targeting both angiogenic and lymphangiogenic vascular changes in tumours.^{[156],[157],[158]} This cost-intensive technique is not applicable to all organs, as especially the liver captures too many phages for a specific selection.^[152]

In addition to proteomic efforts, vascular biomarkers were also explored in transcriptomic studies. Global transcriptomic studies suffer from the high tissue complexity, therefore the analysis is performed on highly specified parts of the isolated tissue, for example via comparison of gene expression patterns from normal and malignant endothelial cells.^{[89],[159]} Nevertheless, the obtained data cannot provide direct information about localization of the corresponding protein target and its

vascular accessibility.

The enrichment of vascular accessible proteins, not only on endothelial cells, but also on the surface of diseased cells or in the perivascular extracellular matrix prior to mass spectrometric analysis heavily facilitates the identification of novel biomarkers. First of all, the data provides direct information about vascular accessibility of the protein target. Furthermore, the enrichment of the surface proteome fraction also enables a more stable identification and quantification from proteins at the lower end of the dynamic range of the proteome, especially membrane proteins.^{[160],[13]} In the following, three experimental approaches for the enrichment of the vascular accessible proteome fraction prior to mass spectrometric analysis are introduced.

1.2.3.1. Carbohydrate-Based Cell Surface Proteome Enrichment

The glycosylation of cell surface proteins plays a pivotal role in various biological processes such as cell adhesion and migration or cell communication.^{[65],[161]} As described in chapter 1.1.2.6, oligosaccharides are hereby attached to proteins through N- or O-glycosidic linkages.^[65] In a malignant state, altered glycosylation patterns are found.^[65] For example, both in hepatocellular carcinoma and in breast cancer, an elevated activity level of N-Acetylglucosaminyltransferase V, an enzyme catalyzing the addition of branched N-linked oligosaccharide side chains, is correlated to increased tumour invasiveness and higher metastasis rates.^{[162],[163]}

One possibility to enrich for glycoproteins is based on the strong, but reversible interaction between lectin and oligosaccharides.^[164] Fractionation of glycopeptides with different oligosaccharide residues is possible by using a series of different lectins with various binding specificities (also termed serial lectin affinity chromatography SLAC).^{[165],[166]} Lectin affinity chromatography was also applied in combination with isotopic labelling (termed isotope-coded glycosylation-site-specific tagging IGOT). The enrichment of glycoproteins is hereby followed by peptide-N-glycosidase (PNGase) catalysed ¹⁸O-incorporation into N-glycosylation sites.^[167] The lectin affinity chromatography methodology is simple and cost-effective. Drawbacks are the possible non-specific interactions of non-glycosylated proteins as well as the poor and often not well-defined selectivity of some lectins.^[161]

Besides their affinity to certain molecules, also the chemical reactivity of the oligosaccharide residues can be exploited for enrichment. Boronic acid derivatives form covalent bonds with molecules containing *cis*-diol groups such as mannose, galactose or glucose.^[168] Magnetic beads functionalised with boronic acid can therefore be used to enrich both for N- or O-linked glycoproteins.^{[168],[161]} Whereas N-glycosylation can be easily removed by PNGase F treatment, O-glycosylation of serine or threonine residues is experimentally more difficult to handle: Thus, a chemical approach based on beta-elimination followed by Michael addition with normal or deuterated DTT (BEMAD) enabling the mass spectrometric analysis of O-linked glycopeptides was developed.^[169] Another tagging approach is based on the enzymatic modification of terminal GlcNAc moieties with a ketone-containing galactose analogue (termed quantitative isotopic and chemoenzymatic tagging (QUIC)).

Subsequently, the ketone can be used for enrichment by chemical modification with aminooxybiotin followed by affinity purification.^[170]

The cell surface capturing (CSC) technique is based on the oxidation of vicinal diol-groups in saccharide residues with periodate. The generated aldehyde groups can be exploited for site-specific coupling to a solid phase such as a hydrazide resin via a hydrazone bond (see Figure 1.3a).^[171] Another approach is coupling of the aldehyde group to bifunctional biocytin hydrazide used for subsequent affinity-based purification.^[172] After enzymatic protein digest, non-glycosylated peptide fragments can be removed by washing, whereas glycosylated peptides remain bound to the resin. Release of N-linked peptides can subsequently be performed with PNGaseF.^[172] As most proteins on the cell surface are glycosylated, the technique was either applied on proteins derived from crude plasma membrane preparations, or applied on living cells upon washing under oxidative conditions.^{[173],[160],[172]} The CSC technique is very specific. In a study on Jurkat cells, 95% of the identified proteins were reported to be cell surface glycoproteins.^[172] In a preliminary study, the technique was also shown to be applicable to whole organs, however requiring the generation of single cell suspensions prior to labelling.^[172] In theory, the technique should be also suitable for the analysis of endothelial cells isolated from diseased or healthy tissue environments.^[160]

1.2.3.2. Silica Coating Technology

The silica coating technology was originally developed for the isolation of plasma membrane proteins of intact cells, and was later adapted for *in vivo* perfusion of rodents.^{[174],[175]} In brief, cell surfaces or the vasculature surface are coated with cationic, colloidal silica binding via ionic interactions. Subsequently, the silica beads are cross-linked with an anionic polymer such as polyacrylic acid, hereby a thin pellicle is formed. After cell lysis or tissue homogenisation, the stabilised and density-increased silica-coated plasma membrane sheets can be purified by density gradient centrifugation. These sheets also include the attached plasma membrane proteins (see Figure 1.3b).^[174]

The methodology was further refined and applied in many mass spectrometric biomarker studies.^{[176],[177]} The enrichment of plasma membrane annotated proteins is reported to be about 50%.^[177] For example, the technique was used for *in vivo* perfusion of the lung vasculature of rats: Comparison to silica coated, *in vitro* cultured rat lung microvascular endothelial cells revealed that 41% of the identified proteins were exclusively identified in the *in vivo* setting.^[93] Analysis of perfused, normal rat lungs vs. lungs of rats bearing metastatic breast adenocarcinoma resulted in the identification of several endothelial biomarkers significantly up-regulated in the tumour model. In the following, it was shown that targeted therapy against Annexin A1 could prolong overall rat survival.^[90]

Limitations of the technique include intracellular contaminations occurring due to attachment of positively charged proteins such as histones to the polyanionic surface after lysis. Furthermore, the degree of the identified intracellular background

proteome fraction is highly sensitive towards the mechanical shear rate used for tissue homogenisation: Harsh conditions are not suitable for the technique, therefore the method is limited to soft tissues.^[178]

1.2.3.3. *In Vivo* and *Ex Vivo* Biotinylation

Another approach for the labelling of the surface proteome fraction is covalent modification with membrane impermeable, reactive ester derivatives of Biotin, which can subsequently be used for affinity-based enrichment of the labelled proteins (see Figure 1.3c).^{[160],[13]} Efficient biotin labelling of plasma membrane proteins on cells could be demonstrated by electron microscopy using streptavidin-gold labelling.^[179] The cell surface proteome labelling technique with subsequent mass spectrometric analysis was applied on various cell types, the identified plasma membrane protein fractions were hereby reported to be 31-67% in different studies.^{[180],[126],[181],[182],[183]}

Vascular perfusion of rodents with biotinylation reagents enables the *in vivo* labelling of the surface of endothelial cells, cells in close proximity of the blood vessels and in the perivascular extracellular matrix. In brief, the animals are terminally perfused with PBS to wash out blood components, followed by perfusion with the reagent and a quenching solution (see Figure 1.3c).^{[184],[185]} The success of the method is dependent on the vascularisation of the target tissue to ensure high biotinylation rates as well as of the efficiency of the biotinylation reagent.^[185] The enriched vascular accessible proteome fractions derived from perfused diseased and healthy mouse organs can be subjected to comparative mass spectrometric analysis to identify novel disease-specific biomarkers for antibody-based therapies.^{[186],[187],[188]} The perfusion technique was also applied *ex vivo* using surgically resected tumour bearing kidney or colon: Perfusion with the biotinylation reagent was performed within 2 min after nephrectomy or ileocelectomy by cannulation of the renal or colic artery.^{[189],[190]} In both studies, tumour-specific biomarker candidates were found (e.g. Periostin in the kidney study). These proteins could additionally be shown by immunohistochemistry to be strongly over-expressed in the tumour stroma.

The linker between the biotin residue for affinity purification and the reactive moiety for protein coupling influences the characteristics and the properties of any biotinylation reagent.^{[191],[12]} In 2010, the novel, peptide-based biotinylation reagent NHS- β -Ala-(L-Asp)₃-biotin was developed, improving the identification of plasma membrane annotated proteins compared to the commercially available reagent Sulfo-NHS-LC-biotin.^[12] The linker design will be further elucidated later in this thesis.

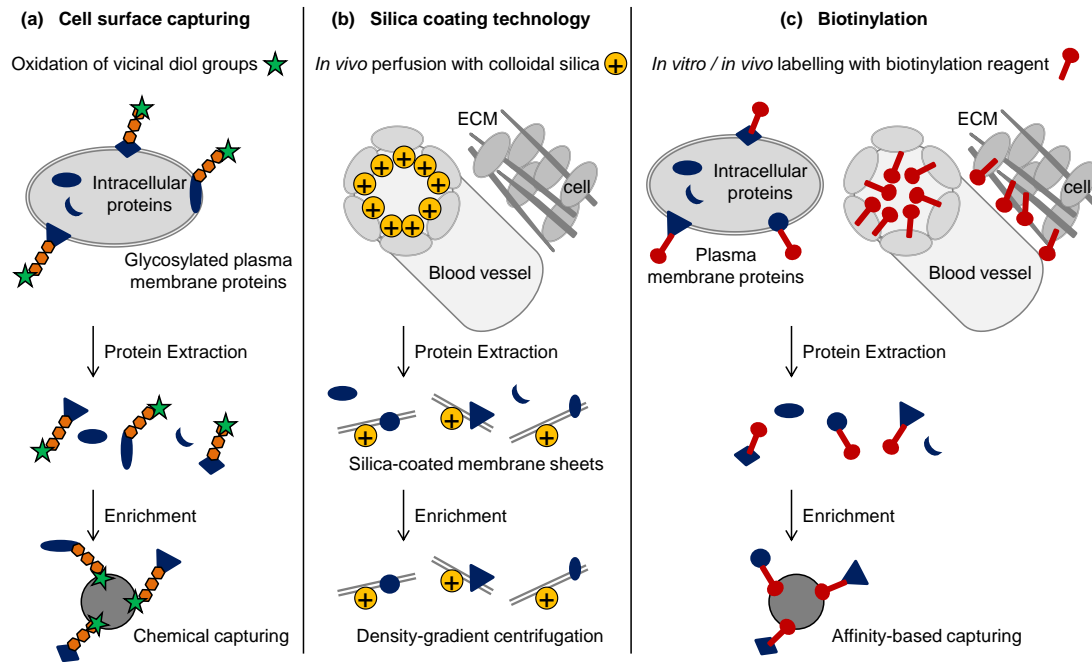


Figure 1.3.: Schematic illustration of three technologies for the enrichment of the surface proteome fraction. (a) Cell surface capturing, (b) Silica coating technology, (c) *In vitro* and *in vivo* biotinylation. See text for further explanations.

1.3. Proteomics and Mass Spectrometry

1.3.1. Mass Spectrometric Approaches

Two basic approaches for mass spectrometric analysis of proteins are known: Top-down proteomics concentrates on the analysis of intact proteins, whereas the bottom-up or shotgun approach is focused on the analysis of enzymatic protein digests. The bottom-up approach is the classical method for large-scale proteomic studies of complex protein mixtures (see Figure 1.4).^[192] Proteins are extracted from their biological source, subjected to enzymatic digest followed by peptide fractionation and mass spectrometric analysis.^[193] In contrast, top-down approaches provide the experimental platform for characterisation of post-translational modifications (PTM) of specific proteins (e.g. complete spatial mapping) or protein isoforms (e.g. alternative splicing forms). A similar in-depth analysis is restricted in shotgun experiments due to the high sample complexity limiting sequence coverage and the assignment of PTMs, the ambiguity of degenerated peptide entries as well as the loss of labile post-translational modifications during sample preparation.^{[194],[195],[196]} Technical challenges in top-down proteomics are the separation and mass spectrometric fragmentation of intact proteins during the analysis, the limited sensitivity as well as the large sample amount needed for analysis.^{[195],[192]} So far, the method works best when applied to simple protein mixtures.^[192] Different attempts to combine

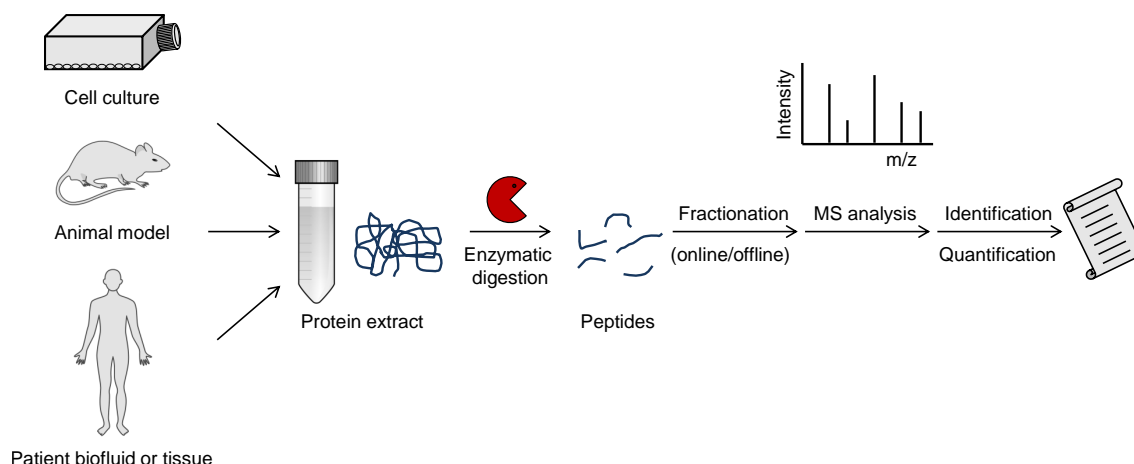


Figure 1.4.: Schematic illustration of a typical shotgun proteomics workflow.

the large-scale protein identification power of bottom-up approaches with in depth-characterisation of single proteins via top-down approaches reach from a compromise of partial digestion with Lys-C into large peptide fragments to combined workflows integrating data from both analysis methods.^{[195],[197]}

The most important sample preparation step in large-scale biomarker discovery studies in proteomics is the reduction of sample complexity.^[198] Efficient separation is the prerequisite for sensitive and accurate measurements as well as for the detection of low-abundant peptides.^[192] Gel-based approaches rely on one- or two-dimensional protein separation via gel electrophoresis and isoelectric focusing followed by gel-band excision, spot picking or slicing of whole lanes (pre-fractionation), and subsequent analysis after in-gel digest.^{[192],[199]} Gel-free approaches are based on protein and/or peptide separation by liquid chromatography (LC). Reversed-phase materials (e.g. C_{18}) are commonly applied to separate peptides according to their hydrophobicity. Ultra-high pressure liquid chromatography systems are capable to operate with small 1.4-3 μm chromatographic particles reported to provide an improved separation efficiency and chromatographic resolution.^[200] Furthermore, multidimensional methods include orthogonal fractionation principles such as strong cation exchange or affinity-based separation.^[192]

Previous enrichment of the target proteome fraction of interest further reduces sample complexity and enables stable identification and quantification, even of low-abundant proteins. Approaches for the enrichment of the vascular accessible surface proteome fraction have already been illustrated in the previous chapter.

Post-translational modifications (PTM) regulate protein structure and function and therefore influence most cellular processes.^[201] Several hundreds of different PTMs are known. Phosphorylation, acetylation, methylation, N- and O-linked glycosylation and ubiquitylation are amongst the most abundant.^{[202],[203]} As PTM deregulation is associated with disease, the global characterisation of PTMs is of greatest interest in proteomics. Hereby, specific enrichment techniques have to

be applied.^[202] The enrichment of glycoproteins was already elucidated in chapter 1.2.3.1. Phosphorylation is crucial as key regulator of many signalling cascades and most commonly occurs at serine residues, but also at threonine or tyrosine side chains.^[204] The use of phosphatase inhibitors is essential during sample preparation to protect the labile modification. Phosphoproteome enrichment is crucial for analysis to compensate for the low stoichiometry of phosphorylated peptides. Enrichment technologies are either based on electrostatic interactions of the negatively charged phosphate group with Fe^{3+} (immobilized metal ion affinity chromatography IMAC) or TiO_2 or on direct or indirect chemical coupling to a solid support by β -elimination of the phosphate group.^{[202],[204]} Other methodologies for PTM enrichment are antibody-based, e.g. the detection of ubiquitylated side chains by antibody recognition of the Lys- ϵ -Gly-Gly motif generated by tryptic digest at former ubiquitylation sites.^[205]

1.3.2. Ion Sources, Mass Analyzers and Protein Identification

In general, a mass spectrometer consists of three parts: the ion source, the mass analyser and the detector to amplify signals and register arriving ions.^[206] Various analysis strategies and instrumental configurations have been developed and are further elucidated in the following.

To introduce the analyte into the system, it is essential to ionize the molecules and to transfer them into the gas phase. For polar organic macromolecules such as peptides or proteins two ionisation methods, both established in the 1980s, are prevailing: electrospray ionization (ESI) and matrix-assisted laser desorption/ionization (MALDI).^{[207],[208],[209]} The development of both soft ionisation methods was crucial for the analysis of biomolecules and was awarded with the Nobel Prize in chemistry in 2002 (John Bennett Fenn, Koichi Tanaka).

ESI is directly compatible with chromatographic separation. Continuous sample separation can be performed using an online system, as the ions are created out of the liquid phase.^[198] The analyte solution is guided through a capillary nozzle whose tip is set under high voltage. Electrostatic Coulomb attraction pulls the ions towards the counter electrode. Hereby, the analyte is dispersed into an aerosol of analyte-solvent droplets.^[210] The solvent in the droplets evaporates, leading to a decrease in droplet size and increase in field density. Surface tension and Coulomb repulsion of like charges are the driving forces for the following declustering process. As soon as the Rayleigh instability limit is reached, Coulomb fission occurs and smaller and more stable droplets are created.^[210] The ion evaporation model and the charge residue model have been hypothesized to further describe the processes.^[210] Ions generated by ESI are often multiply charged due to charge redistribution within the droplets formed.^[210]

The use of MALDI uncouples the chromatographic separation from the analysis, which also enables uncoupling of MS1 and MS/MS analysis as well as reanalysis of the sample.^[198] The analyte is mixed with a matrix excess and manually or robotically co-crystallized as spots on a MALDI target plate. Different matrices

are available for various application: α -Cyano-4-hydroxycinnamic acid (CHCA) and 2,5-dihydroxybenzoic acid (DHB) are commonly used for peptide analysis, Sinapinic acid for protein measurements.^[211] The organic matrix system is able to absorb energy from intense laser pulses at a defined wavelength, energetic relaxation within the crystal lattice leads to explosive desorption of matrix molecules together with the analyte molecules.^[212] Proton transfer from the photo-excitated matrix molecules to the analyte ($[X+nH]^{n+}$) is hypothesized as ionisation mechanism, hereby mostly singly charged molecules are formed.^[212] The ionisation technique currently gains attention by the development of imaging mass spectrometry, enabling the spatial analysis of many molecules on clinical tissue sections coated with MALDI matrix.^{[213],[214]}

The first type of mass analysers were magnetic or electric sector field instruments. The trajectories of the ions within the system are determined by their mass-to-charge ratio (m/z). Differences in the deflection within the magnetic or electric field are detected.^[215] Time-of-flight (TOF) analyzers accelerate the ions in an electric field, followed by a field-free drift tube under high vacuum. According to the kinetic energy and their specific m/z ratio, the ions are separated by different flight velocities.^{[206],[199]} The introduction of an reflector as an ion mirror not only doubles the flight path, but also corrects for slight differences in the kinetic energy (focusing).^[206] An integrated collision cell in TOF/TOF instruments enables peptide fragmentation by collision-induced dissociation (CID).^[206] Quadrupole analysers are based on the stabilisation or destabilisation of the path of an ion of a specific m/z in a time-varying electric field between the four rods. In typical triple quadrupole instruments, a precursor ion scan is performed (MS1), then, guided by the detected MS1 signal intensity, Q_1 submits a selected m/z to Q_2 , acting as a collision cell and inducing fragmentation, followed by fragment ion scan (MS/MS) in Q_3 (see Figure 1.5). Mass accuracy and resolution of triple quadrupole instruments are lower

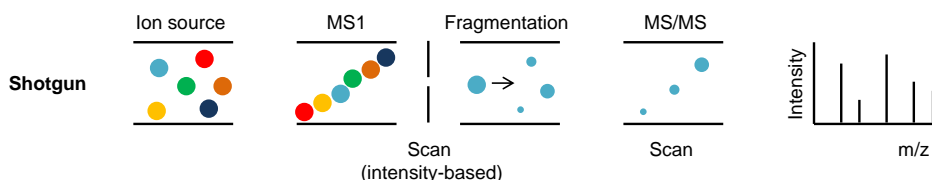


Figure 1.5.: Schematic illustration of a mass spectrometric shotgun analysis on a triple quadrupole instrument.

than for the TOF instruments, whereas the dynamic range and the detection limit are increased.^[216] Ion trap analysers are based on the trapping of ions followed by sequential ejection and scan of ions of a specific m/z . Linear ion traps are constructed as quadrupole, whereas non-linear ion traps consist of two hyperbolic electrodes with a ring electrode in between.^{[206],[199]} Whereas the sensitivity is high, resolving power and accuracy are limited.^{[199],[216]} Fourier transform ion cyclotron resonance analysers (FTICR) and Orbitrap mass analysers are based on a fast Fourier transform. In Orbitrap analysers ions oscillate in circuits around a spindle-shaped electrode. The time-domain signal is mathematically transformed to a frequency-domain signal,

which is directly linked to the m/z ratio.^{[217],[218]} Fourier transform based analysers provide high mass resolution and accuracy.^{[216],[218]} Hybrid instruments are developed to integrate various performance characteristics in one mass spectrometer.^[215]

As already illustrated in Figure 1.5, mass spectrometric protein identification in shotgun experiments is based on a survey precursor scan (MS1), followed by intensity-based selection of 1-50 precursors, that are fragmented and analysed (fragment ion scan, MS/MS).^{[219],[220]} The data gained from each MS/MS spectrum can be directly linked to a corresponding precursor peptide m/z .^[221] The most common way for peptide fragmentation is collision-induced dissociation (CID). Backbone fragmentation hereby occurs upon collision with inert gas molecules such as N_2 , typical fragment spectra of tryptically digested and CID fragmented peptides are dominated by the y- and b-ion series (see Figure 1.6).^[222] Other fragmentation techniques are electron capture dissociation (ECD) or electron transfer dissociation (ETD), which are based on the direct (ECD) or indirect (ETD) transfer of an electron to the peptide ion. The instability of the formed radical subsequently leads to backbone fragmentation, preferentially leading to c- and z-ions (see Figure 1.6).^[222]

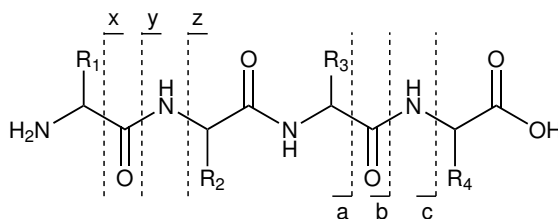


Figure 1.6.: Nomenclature of peptide fragment ions.

MS/MS spectra are subsequently searched against a database using the *in silico* digested protein sequences to identify peptides and the corresponding proteins.^[219] Today, there are many search engines and algorithms known to handle uninterpreted MS/MS data.^[221] Search parameters, including a sequence database, the digestion enzyme used, introduced modifications (e.g. alkylation of cysteine residues) as well as the mass tolerance (depending on the instrument type used) are variable and depending on the experiment performed.^{[221],[219]} Various scoring algorithms are known to score peptide matches, e.g. based on probability such as in Mascot.^[221] Besides, a false discovery rate (FDR) is calculated to validate the search results. A common way is a repeated search against a decoy database with reversed or shuffled sequences. The number of matches against the decoy database is an estimate of the percentage of false positive hits out of the search against the target database.^[221] The identified peptide sequences are finally assigned to the corresponding proteins. The challenge of handling degenerated peptide entries and proteotypicity is further elucidated later in this thesis.

1.3.3. Quantification

Quantitative proteomics is usually dealing with the ratio-based quantification of protein up- or down-regulation between different samples (relative quantification). In contrast, absolute quantification presents exact amounts of the protein present, nevertheless the quantification is relative to an internal standard such as the corresponding peptides synthesized with a heavy label.^[223] The targeted SRM approach for the quantification of several protein candidates in one run is presented in chapter 1.4.1. The quantification of large-scale shotgun proteomics data, either label-based or label-free, is subject of this chapter.

1.3.3.1. Label-Based Quantification

Three basic approaches have been used to integrate labels into peptides and proteins for quantification: Metabolic, enzymatic and chemical labelling. A variety of techniques differing in the point in time of label introduction and the method used for quantification have been developed. In the following, some of the most frequently used techniques and reagents are presented.

Stable isotope labelling with amino acids in cell culture (SILAC) is based on metabolic labelling of cultured cells with isotopic heavy or light versions of amino acid. For combination with standard tryptic or Lys-C digestion, typically $^{13}\text{C}_6$ $^{15}\text{N}_2$ -lysine and $^{13}\text{C}_6$ $^{15}\text{N}_4$ -arginine are used for heavy labelling, generating a mass difference of 8.0142 Da and 10.0083 Da, respectively.^[224] Light and heavy labelled protein extracts are mixed in a ratio of 1:1, digested and analysed. Relative quantification is based on the ratio of the corresponding peak intensities of the light and heavy form in MS1.^[224] To enable relative comparison of tissues samples, the super-SILAC strategy was developed: Several cell lines representing the corresponding tissue type are grown in heavy medium and combined to a heavy super-SILAC mix, which is subsequently used for relative quantification against the light proteins derived from tissues samples.^[224] Furthermore, the SILAC method was also applied *in vivo*. Metabolic labelling of *C. elegans* or *D. melanogaster* could be achieved by feeding ^{15}N -labelled *E. coli* or yeast.^[225] Moreover, SILAC mice could be obtained with a $^{13}\text{C}_6$ -lysine containing diet. Incorporation of the heavy label is approximately 85% within 8-10 weeks, the F2 generation is reported to be fully labelled.^[226]

Proteolytic ^{18}O labelling is an enzymatic method to introduce a heavy isotopic modification. The labelling hereby occurs via enzymatic digest in heavy water (H_2^{18}O), leading to the incorporation of ^{18}O into the peptides' carboxyl terminus.^[227] Light and heavy samples are combined after the enzymatic digestion. Peptides are distinguished by a 2-4 Da mass shift, depending on the introduction of one or two ^{18}O atoms.^[227] Main disadvantage of the method is the possible trypsin-mediated back-exchange of ^{18}O to ^{16}O occurring in the aqueous solvents used. Furthermore, mass differences of less than 4 Da complicate accurate quantification.^[223]

Chemical labelling for quantification is either performed by modification with isotope-coded tags or with isobaric tags. The isotope coded affinity tag (ICAT)

approach was one of the first labelling methods introduced for quantification.^[228] The reagents contain a thiol specific functionality targeting protein cysteine residues, a biotin residue to purify the labelled peptides after tryptic digest and a light or heavy (8x deuterated) linker, to evaluate the ratios of the corresponding peak intensities of the light and heavy samples in MS1.^[228] The reduction of sample complexity by enrichment of the relatively rare cysteine containing peptides is at the same time the weakness of the method, as the quantification is based on a low number of peptides per protein.^[223] Many variants of the heavy isotope tags have been developed. For example, the isotope coded protein label (ICPL) reagents specifically modify free amino groups with a heavy or light nicotinyl-N-hydroxysuccinimide derivative.^[229] The phosphoprotein isotope-coded affinity tag (PhIAT) approach specifically targets O-phosphorylation sites with a heavy or light labelled affinity tag.^[230]

The isobaric tag for relative and absolute quantification (iTRAQ) approach is based on the modification of free amino groups (lysine side chains or peptide N-termini) of digested peptides. The reagents are composed of a reporter group differentially labelled with ^{15}N and/or ^{13}C (with the 4-plex reagents 114-117 Da) and a balancer group (with the 4-plex reagents 191-188 Da), assembled to identical reagent masses. The labelled peptides are therefore indistinguishable in the MS1 spectra. Upon MS/MS fragmentation, the reporter group is cleaved and peptide abundance differences can be quantified based on the reporter group signal intensity.^[231] So far, 4- and 8-plex for multiplexing four or eight samples within one measurement are known.^{[231],[232]} Variants of isobaric tags are the 6-plex tandem mass tag (TMT) reagents, also targeting free amino groups, and the cleavable isobaric labelled affinity tag (CILAT), allowing for 12-plex measurements, which is the maximum known so far.^{[223],[233],[234]}

The main advantages of chemical labelling is that the compared samples are processed and measured in one run, which minimizes variability. Besides the high reagent costs, limitations of labelling technologies are the variable labelling efficiency and the additional sample preparation steps necessary for introduction of the label resulting in decreased reproducibility.^[234] Furthermore, sample complexity is increased by multiplexing approaches and the detectable dynamic range can be as low as three orders of magnitude.^{[223],[234]} Moreover, the number of comparisons is limited by the multiplexing capacity of the applied reagent tag.^[234]

1.3.3.2. Label-Free Quantification

Label-free quantification is purely based on data analysis. Theoretically unlimited sample numbers can be processed in individual measurements without any additional labelling steps.^[234] The increased sensibility of label-free methods for handling errors or technical performance differences emphasizes the need for stringent control of data quality as well as for data normalisation.^{[234],[223]}

The simplest estimation of protein abundance can be calculated by the exponentially modified protein abundance index (emPAI). The protein abundance index PAI is defined as the ratio of the number of observed peptides divided by the num-

ber of calculated observable peptides by *in silico* digest under a given mass range detectable by the mass spectrometric method used.^[235] The emPAI value is calculated as $10^{\text{PAI}-1}$ and was shown to correlate with the protein amount within a sample ($r = 0.89$).^[236] The absolute protein content in mol% can be estimated as percentage of the emPAI value of a specific protein divided by the sum of all emPAI values.^[236]

Label-free quantification approaches are dominated by two different principles: Quantification is either based on the area under the curve (AUC) derived from the precursor ion spectra (MS1, signal intensity based) or on spectral counting, based on the assumption that more abundant peptides will generate more MS/MS spectra (see Figure 1.7).^{[234],[237]}

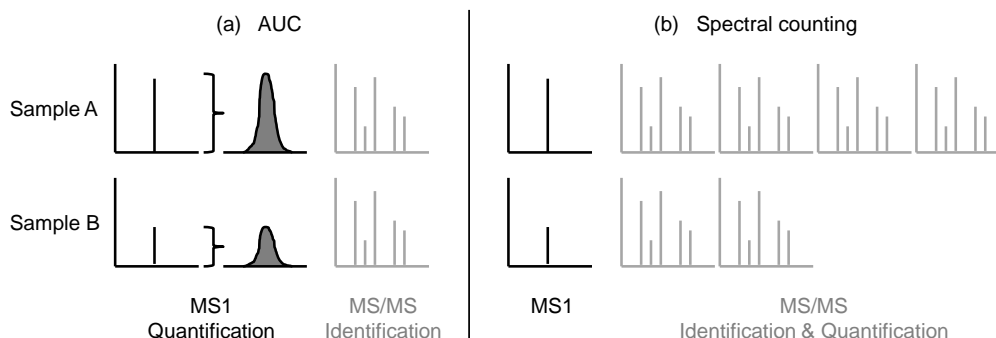


Figure 1.7.: Schematic illustration of the two main principles of label-free quantification. Quantification based on the area under the curve (a) or on spectral counting (b).

Spectral counting assumes that high-abundant peptides are present in more chromatographic fractions and are more often selected for MS/MS analysis, therefore more spectra are produced.^[237] The method has undergone several modifications to include normalisation factors and to improve accuracy. Usage of the normalised spectral abundance factor (NSAF) corrects each spectral count for the protein length, as larger proteins produce more peptides and therefore a higher number of spectra than smaller ones. The value is subsequently normalised against the sum of all spectral count per protein length ratios from all proteins within the experiment.^[238] The APEX (absolute protein expression) approach adds a correction factor to spectral counting, which is based on peptide detection probabilities derived from a machine learning classification algorithm that takes into account different peptide characteristics such as mass, sequence, length and hydrophobicity.^[239] Main disadvantage of spectral counting is the limited dynamic range of only two orders of magnitude. Saturation of the detector occurs for higher abundant peptides and signal suppression for lower abundant ones, hindering a linear correlation.^[223] Spectral counting is barely used combined with MALDI as ionisation method. A high number of MS/MS spectra has to be recorded to obtain a stable quantification, whereas a specific peptide is found only in a few spots and maximally recorded once per spot on a MALDI target plate.^[234]

Ion intensity based quantification relies on MS1 data. Extracted ion chromatograms (XIC) are generated by grouping together mass spectrometric signals derived

from the same peptide, followed by integration over the chromatographic scale.^[240] Linear correlation of the area under the curve (AUC) to the protein concentration could be shown in a range of 10 fmol-100 pmol ($R^2 = 0.991$).^[241] The prerequisites of a successful signal intensity based quantification are reproducible LC/MS runs to enable correct feature alignment across the different samples using chromatographic retention time and m/z ratio. Slight and unavoidable run-to-run variability emphasizes the need for data normalisation.^[242] In general, data processing is performed in the following steps: (i) After smoothening of the raw spectra and baseline subtraction, the peptide peaks are detected. (ii) Data from detected isotopic patterns are merged in a deconvolution step. (iii) Chromatographic retention time is aligned and the peak intensity is normalised. (iv) Detected peaks deriving from the same peptide (elution and measurement in several fractions) are summed to so-called features based on retention time and m/z . For ratio-based quantification across different samples, corresponding features are matched, also based on retention time and m/z .^{[242],[234]} The identification information derived from the MS/MS spectra is assigned to the quantification information during the workflow. Most pipelines are summarization-based, i.e. the intensities of peptides derived from one protein are finally summed for analysis.^{[234],[243]} Recently, a peptide-based linear regression model was proposed correcting for peptide characteristics (e.g. varying signal intensity derived from different ionisation efficiency) and the different number of identified peptides per protein.^[243]

1.3.4. Data-Independent Acquisition

The shotgun proteomic discovery workflow presented in the previous sections represent the so-called data-dependent acquisition (DDA) method. Thereby, the survey or precursor ion scan (MS1) is the basis for intensity-based selection of precursor ions subsequently submitted to MS/MS analysis (see Figure 1.5). In a typical mass spectrometric measurement 1-50 precursor ions are selected for fragmentation and sequential analysis. Precursor and the corresponding fragment ions can hereby be clearly assigned.^[220] Efficient complexity reduction is of greatest importance to enable the measurement of low-abundant peptides exhibiting a decreased signal intensity.^[220] Disadvantages of DDA include the potential under-sampling, especially of low-abundant peptides, the limited reproducibility of the measurements as well as analytical problems arising from co-fragmentations of peptides.^[244] Targeted proteomic approaches such as SRM (see chapter 1.4.1 and Figure 1.9) investigate specific fragment ions from defined target peptides and provide a highly reproducible quantification, but are restricted to less than 100 proteins per run.^[220]

The idea behind the data-independent acquisition (DIA) mode is the unbiased fragmentation of all precursors followed by acquisition of all product ion spectra. Thereby, a high number of analytes can be detected under high sequence coverage, accuracy and reproducibility.^[220] Recent developments in mass spectrometry created diverse approaches based on the DIA principle.

Classical sequential window acquisition of all theoretical mass spectra (SWATH)

analysis is based on cycling through 32 sequential precursor mass windows of 25 Da width, followed by fragmentation of all precursors within the predefined m/z and within the specific retention time range under acquisition of all product ions (see Figure 1.8).^[220] As the spectra are multiplexed and fragment ions cannot be clearly

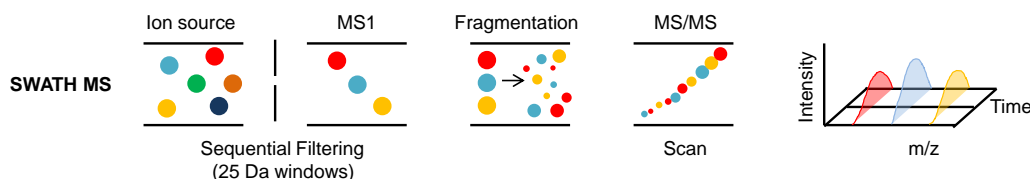


Figure 1.8.: Schematic illustration of a mass spectrometric SWATH analysis on a Triple TOF instrument.

assigned to a precursor, a simple database search is not compatible with this type of data. Powerful bioinformatic tools assign the DIA data to spectral reference maps (assay libraries) containing peptide coordinates such as precursor and fragment m/z , normalized precursor retention times and relative intensities of fragment ion signals.^[245] The reference libraries are generated from data acquired in shotgun proteomics experiments (DDA analysis) and by sequencing of libraries of synthetic peptides.^[220]

Multiplexed MS/MS (MSX)-DIA uses five randomly selected 4 Da windows distributed over the whole mass range per scan (in total 20 Da window). Maximal 10% of 4 Da windows are reported to contain more than one peptide.^[244] Thereby, the precursor sensitivity is increased, as the whole mass range is exploited and the simultaneous isolation of peptides and modified forms of those peptides (e.g oxidized methionine) generating overlapping fragment spectra is prevented.^[246] Furthermore, SWATH acquisition variants with variable cycle times or variable window widths to adapt to the present peptide density is reported.^[247] An optimized DIA method tested on a yeast digest was able to provide almost 4000 protein identifications in a one hour run with good reproducibility (83% overlap between five replicate samples).^[248]

The MS^E approach exploits rapid switches in the collision cell between low and high collision energies. First, at the low energy state, the precursor spectrum is obtained. Subsequently, the high energy state favours fragmentation. MS/MS spectra are recorded, hereby the collision energy is ramped.^[244] The high-definition MS^E (HD MS^E) technology includes travelling wave-based ion mobility separation (IMS) as additional separation dimension to expand the system's peak capacity.^[249] During the high-energy scan, a fixed collision energy is set to each IMS cycle. With HD MS^E , approximately 50% more proteins could be detected in comparison to the MS^E approach.^[249] In early IMC cycles with low collision energy, peptides with high m/z remain intact, whereas they are efficiently fragmented in late IMS cycles using high collision energy, while low m/z are over-fragmented. To optimize the measurement efficiency by increasing the scan time spent within favourable collision energy windows, ultra-definition MS^E (UD MS^E) was suggested. The technique further exploits

the linear correlation between m/z and the collision energy as well as the correlation of ion mobility with the m/z ratio. The collision energy is ramped within a specific range during each IMS cycle, late IMS cycles are hereby assigned to a range of higher collision energies.^[249] UDMS^E led to an additional increase of protein identification rates, i.e. approximately 50% more proteins were identified compared to HDMS^E.^[249]

The idea behind data-independent acquisition is to miss less information in a mass spectrometric run compared to DDA or SRM approaches. A future concept is to perform high-throughput experiments of clinical samples and to store their complete proteomic profiles. The profiles can be re-analysed at any point in time to answer new clinical questions or to validate novel biomarker candidates.^[220] Future developments not only include technical developments, but also the establishment of better associated bioinformatic tools for data analysis and reference libraries.^[220] Furthermore, instrumental improvements in scan speed and high resolution will significantly improve and facilitate both data-independent and data-dependent analysis.

1.4. Biomarker Validation and Targeted Therapies

High-throughput proteomics experiments generate a large number of highly regulated protein hits, yet, only very few biomarker candidates are still successful in the post-discovery phase and are finally investigated in clinical studies.^[129]

First key point is the study design itself.^[9] The access to clinical material is often limited and the analysis time consumptive, therefore often limited sample numbers are analysed. Results obtained from small sample cohorts might be over-interpreted. Detailed information about the representative cohort are essential for further classification e.g. by the stage of disease, treatment obtained and histopathological characteristics (subtypes). Case and control specimens have to be matched based on several characteristic such as age and sex to obtain a valuable comparison. Standard procedures in sample collection, storage and handling minimize artificially introduced variability. Technical limitations such as slight variability in sample preparation as well as reproducibility of the measurements emphasize the need for technical replicates to enable statistical assessment.^{[9],[250]}

Secondly, validation of the biomarker is crucial to evaluate its performance (e.g. sensitivity and specificity characteristics) in a larger sample cohort and to identify the technical assay conditions yielding in highly reproducible and precise measurements.^{[251],[250]} Several validation techniques typically used for protein biomarkers are further elucidated in chapter 1.4.1.

Thirdly, to exploit a biomarker for targeted therapy, binding antibodies have to been developed (e.g. by phage display), which is time and cost intensive and requires excessive follow-up validation experiments.^[252] The *in vivo* situation mostly assessed by use of animal models provides information about accessibility of the target as well as biodistribution and associated side-effects. Furthermore, blood clearance characteristics of the drug as well as toxicological safety have to be determined.^[252]

1.4.1. Validation of Protein Biomarkers

In contrast to the discovery experiments, typical validation experiments are based on a decreased analyte number (biomarker candidate panel), whereas sample numbers are increased.^[253] Traditionally, antibody-based methods such as ELISA on body fluids or protein extracts and immunohistochemical stainings of tissue specimens are exploited.^[220] Increased sample throughput is enabled by screening multiple tissue sections on tissue microarrays.^[46] A high quality antibody is capable to provide high sensitivity and selectivity characteristics. Nevertheless, antibody-based methods have many limitations. Cross-reactivity often limits selectivity and/or specificity of an antibody, depending on the antibody's affinity and the concentration of the cross reacting substance within the sample.^[45] Multiplexing is difficult and limited due to cross-reactions between the different antibodies applied. Furthermore, tissue fixation and preservation affects the staining procedure. Optimization of the sample preparation and evaluation of different antibodies is crucial, but time and cost intensive.^[220] Though, for many novel candidate proteins, good quality antibodies are not available so far.^{[220],[253]}

Mass spectrometric selected reaction monitoring (SRM) based quantification provides high reproducibility, accuracy and sensitivity, a dynamic range of up to five orders of magnitude and multiplexing capability.^{[220],[253],[250],[254]} Up to 100 proteins can be targeted in one run.^[220] The SRM analysis is usually performed on a triple quadrupole instrument with Q_1 as precursor mass filter and Q_3 as fragment ion mass filter (see Figure 1.9). The precursor ion of a predefined target peptide is selected,

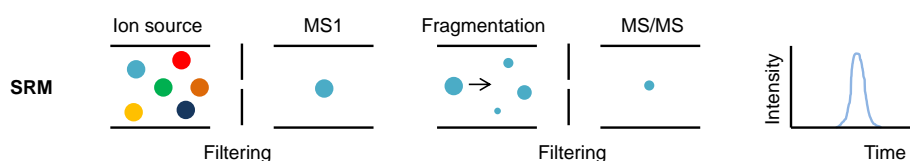


Figure 1.9.: Schematic illustration of a mass spectrometric SRM analysis on a triple quadrupole instrument.

fragmented in Q_2 and a predefined fragment ion type, the so-called transition, is guided to the detector. The SRM signal acquired is derived from monitoring the transition over a predefined retention time window.^[220] The prerequisite of a SRM analysis is the selection of proteotypic peptides unambiguously identifying the target proteins within the sample. Secondly, the targeted fragment ions (transitions) are selected during optimisation experiments, mainly based on their signal intensity.^[223] Synthetic heavy labelled peptides spiked into the sample provide information about mass spectrometric performance differences and enable quantification via the ratio of the peak area or height of the light sample peak and the heavy standard peaks.^[223] Ion suppression effects are negligible for the quantification, as a similar behaviour of sample and standard peak can be expected.^[223] Still, the analysis of low-abundant peptides might be hindered by high-abundant peptides in complex

samples. Furthermore, the injectable amount is limited due to technical restrictions of the LC/MS system.^[223]

Combinations of antibody-based enrichment methods with mass spectrometric quantification are reported, to enrich for the target proteins prior to analysis enabling a more stable quantification.^[223] In the SISCAPA method, specific anti-target peptide antibodies are immobilised on a solid support, which is subsequently packed into an affinity column. The eluted fractions are guided via a reversed-phase column to SRM analysis.^[255] The SRM analysis of plasma proteins, spanning a dynamic range of more than ten orders of magnitude, could be heavily improved by SISCAPA. About 120-fold enrichment could be achieved for over 200 target components.^[255] Successful enrichment also yielded in an improved limit of detection (LOD). The LOD for SRM analysis of unfractionated plasma samples is about 1 µg/mL. SISCAPA improved the LOD to the low ng/mL range.^[250]

1.4.2. Ligands for Vascular Targeting Applications

In contrast to conventional therapeutics, targeted drugs preferentially localise at the site of disease, preventing side-effects in healthy tissue and increasing the dose finally reaching the diseased tissue.^[252] Besides monoclonal antibodies, also other high affinity binders are known, which are able to target vascular accessible biomarkers.^[91]

For some targets, the isolation of short peptide ligands is possible.^[91] The synthesis and selection of peptide binders are enabled by the phage display technique using peptide libraries presented on the bacteriophages' surface or by SPOT synthesis allowing the generation of a large peptide pool in a heavily parallelised approach.^{[256],[257]} Dissociation constants for small peptide binders are usually in the micromolar range, but can be improved by peptide multimerization to reach the nanomolar range.^{[91],[258],[259]}

Aptamers typically exhibit dissociation constants in the micromolar to subnanomolar range.^[91] Target-binding nucleotide sequences can be selected *in vitro* out of a random pool of single and double stranded DNA sequences in cycles of selection and amplification.^{[260],[261]} Single-stranded DNA aptamers are able to create a larger variability of three-dimensional shapes and exhibit increased affinity and specificity.^[260] The *in vivo* stability can be improved by creating biostable compounds, e.g. so-called RNA-Spiegelmers by exchanging the D-ribose backbone with its nuclease-stable L-form.^[262] Furthermore, bioorthogonal conjugation of drug molecules to aptamers for targeted drug delivery has been performed.^[263] In comparison to antibodies, the main disadvantage of both aptamers and small peptide ligands is the increased clearance rate due to the small molecule size.^[263]

Isolation of high-affinity small organic molecules for vascular targeting provides many advantages. The production is relatively simple, chemical modifications can easily be introduced, small molecules are not immunogenic, oral administration and a good distribution within tissues is achievable due to the small molecular weight.^[91] Several methods to provide small molecule libraries are known. Libraries of synthetic organic molecules used for screening are often created by modular variation of side

groups of compounds inspired by naturally occurring small molecules.^[264] Combinatorial chemistry is based on the assembly of several building blocks to many diverse molecules using a defined number of functional groups. Dynamic combinatorial chemistry hereby exploits reversible reactions, enabling compound rearrangement during the selection process on the target molecule.^[265] Encoded self-assembling chemical libraries are based on libraries of small organic molecule compounds that are coupled to defined oligonucleotide sequences used for decoding after target binding.^[266] Dissociation constants can extend to the nanomolar range, but the isolation of high-affinity binders is still challenging.^[91] Multivalent binding hereby helps to compensate for lower affinity.^[267] So far, several small organic ligands for tumour targeting have been discovered, e.g. a folate derivative targeting the highly up-regulated folate receptor in ovarian cancer, which is currently investigated in a phase II study using a ^{99m}Tc-coupled derivative for imaging purposes.^{[267],[268]} Also coupling to cytotoxic payloads can be easily realised with small organic molecule.^[267]

For efficient targeting, dissociation constants of at least 10 nM are reported.^[267] High affinity antibodies can reach dissociation constants down to the picomolar range.^[91] Different antibody formats are known, reaching from full immunoglobulins and bispecific antibodies to single chain fragments and diabodies.^{[91],[73],[269]} The antibody format heavily influences tissue penetration and pharmacokinetic properties. Biological half-life is increased if the antibody mass is above the renal clearance threshold of 70 kDa. Molecular size, shape, charge and affinity influence the tissue penetration capability. An 80% fraction of monoclonal antibodies (about 150 kDa) is reported to not leave the blood and reach the diseased tissue.^[73] Traditionally, monoclonal antibodies are prepared by the hybridoma technology via the fusion of myeloma cells to spleen cells of an immunised donor.^[270] Newer technologies for synthesis and selection of high affinity antibodies or antibody fragments include phage display or ribosome display as well as colony filter screen.^{[271],[272],[273]} Monoclonal antibody based therapeutics can affect diseased cells in various ways such as receptor blockade interfering with signalling pathways or checkpoints, inhibition of angiogenesis or via co-activation of immune effector cells.^[274] Furthermore, antibodies can be coupled to a payload for diagnostic or therapeutic purposes. Possible payloads can hereby be radionuclides, cytotoxic agents, immunocytokines and nanocarriers.^{[275],[276],[277],[267],[278]} The success of antibody-drug conjugates is not only influenced by the antibody properties, but also by the kinetics of drug release and the diffusion characteristics of the payload within the diseased tissue.^[267] Site-specific attachment of the payload to the protein antibody can be performed via bioorthogonal coupling methods. Cleavable linkers can efficiently release the payload at the targeted site.^[267] Hereby, diverse strategies such as acide-labile hydrazone linkers, disulfide linkers sensitive to a reductive environment or enzymatically cleavable peptide-based linkers can be applied.^[279] Currently, several clinical studies are performed on non-internalizing antibody-drug conjugates delivering their payload to the extracellular space, for example armed antibodies targeting alternative splice variants of the EDA domain of fibronectin.^{[267],[15]}

The development of novel targeted therapies is first of all dependent on the identi-

fication of vascular accessible biomarkers. The use of state-of-the-art analytical technologies, the accessibility of a broad range of clinical specimens and an elaborated biomarker discovery and validation pipeline might be the key factors for successful discovery experiments. Equally important is the ligand development phase, which does both involve production of a binder and extensive validation and assessment of its affinity, pharmacokinetic and toxicological characteristics.

2. Aim of the Thesis

The development of novel targeted therapies is based on the identification of vascular accessible, disease-specific biomarkers. The molecules of interest can be targeted with monoclonal antibodies, antibody fragments or antibody-drug conjugates and are located on the surface of diseased cells, on newly formed blood vessels or in the perivascular extracellular matrix. Proteomic approaches for biomarker identification are challenging due to the enormous dynamic range of the proteome over at least seven orders of magnitude, the low abundance of the highly diverse plasma membrane proteins, as well as the hydrophobic character of membrane proteins. Therefore, the targetable surface proteome fraction is often under-represented in full proteome datasets or cannot be quantified stably. The proteins of interest can be enriched prior to mass spectrometric analysis by their covalent modification with a membrane-impermeable ester-derivative of biotin, followed by streptavidin-based affinity capturing. Biotinylation of potential biomarkers can be conducted *in vitro* by labelling of cells, via *in vivo* perfusion of mice and *ex vivo* using surgically resected tissue material. The linker between the biotin residue and the reactive group for protein coupling hereby determines the properties of the biotinylation reagent and influences selectivity and reactivity.

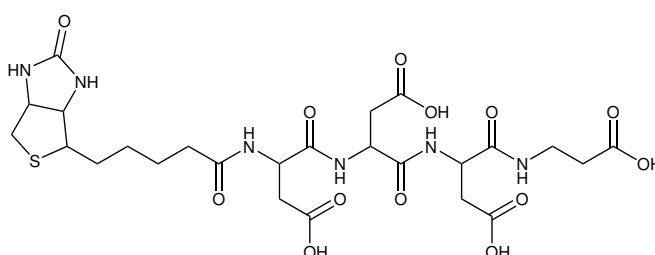
The work in this thesis focuses on the synthesis and multi-step validation of two novel, peptide-based and site-specifically activated biotinylation reagents as well as of two novel heparin-based biotinylation reagents. Furthermore, two alkyne-tagged reagents for bioorthogonal click-chemistry based enrichment are designed and preliminarily tested for reactivity. First, the reactivity is examined by coupling to a single protein in different ratios and analysis with linear MALDI mass spectrometry. Subsequent validation is performed *in vitro* on HeLa cells and *in vivo* via perfusion of healthy NSG mice. The reactivity as well as the efficacy of biotinylation and the *in vivo* biotinylation depth around blood vessels is hereby visualised via confocal laser scanning microscopy, FACS analysis, ELISA and Western Blot of cell and tissue samples. Main part of the analysis is the mass spectrometric determination of the accessible proteome fraction performed in comparison to the commercial reagents Sulfo-NHS-LC-biotin and NHS-PEG₁₂-biotin as well as PBS- or non-treated cells or tissue samples as negative controls. Data analysis concentrates on the identified protein and proteotypic peptide numbers as well as on protein localization, relative quantification is performed against commercial Sulfo-NHS-LC-biotin using the in-house developed software MS_QBAT. Furthermore, the stability of the quantification of some medium- and low-abundant cell surface and extracellular matrix proteins is further assessed by SRM-analysis in comparison to full proteome samples.

3. Materials and Methods

3.1. Chemical Synthesis

3.1.1. Peptide-Based Biotinylation Reagents

3.1.1.1. Synthesis of Biotin-(L-Asp)₃-β-Ala

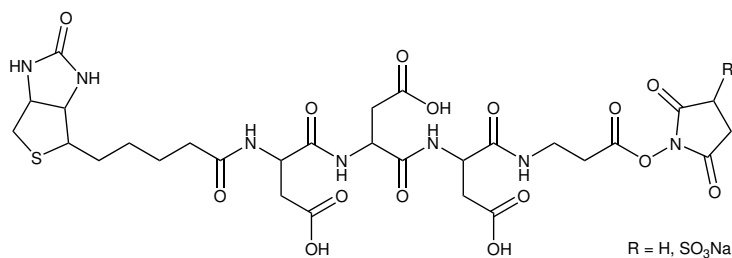


1

Biotin-(L-Asp)₃-β-Ala **1** was kindly provided by Dr. Yixin Zhang (B Cube, Dresden). Peptide synthesis was performed by solid-phase synthesis on an Intavis peptide synthesizer using 9-fluorenylmethoxycarbonyl (Fmoc) chemistry based on a β-Ala-Wang resin (0.55 mmol/g). Activation of Fmoc-Asp-t-butyl-OH was achieved with 0.5 M HBTU in dimethylformamide (DMF). N-terminal deprotection was performed in 20% piperidine in DMF. Cleavage from the resin and deprotection of aspartic acid side chains was achieved with a mixture of 92.5% trifluoroacetic acid (TFA), 5% triisopropylsilane (TIPS) and 2.5% water for 2.5 h at room temperature. After precipitation in cold anhydrous diethyl ether, the peptide was dried under vacuum, followed by preparative reversed-phase HPLC (Luna[®] C18 column, 10 μm particle size, 30 x 250 mm, Phenomenex) with a 30 min linear water/acetonitrile gradient (0.1% TFA, 20 mL/min) for purification. For monitoring purposes, wavelengths of the UV/VIS diode array detector (Varian ProStar 325) were set to a range of 214–280 nm. Biotin-(L-Asp)₃-β-Ala **1** was lyophilized and the purity was verified by an analytical Waters UPLC/MS system (Acquity BEH C18 column, 1.7 μm particle size, 2.1 x 50 mm, Waters) with tandem quadrupole (TQ) and tunable UV (TUV) detector. Two synthesis batches of 25 mg peptide product were used for the experiments presented in this thesis.

Biotin-(L-Asp)₃-β-Ala **1** (C₂₅H₃₆N₆O₁₃S₁): MW 660.65 g/mol.

3.1.1.2. Synthesis of Biotin-(L-Asp)₃-β-Ala-NHS and -sNHS



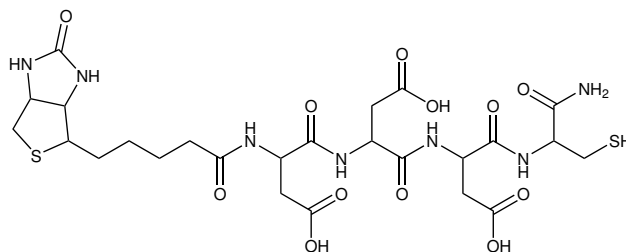
2,3

Biotin-(L-Asp)₃-β-Ala **1** was dissolved in dry dimethyl sulfoxide (DMSO) (over molecular sieve, Sigma-Aldrich, 41648) to a final concentration of 200 mg/mL (303 mM). The stock solution was stored under inert gas at -20°C for several months. For carboxylic group activation, 1.3 eq (1.3 μmol, 150 μg) N-hydroxysuccinimide (NHS) (200 mM in dry DMSO, ProteoChem, c1101) or 1.3 eq N-hydroxysulfosuccinimide (sNHS) (1.3 μmol, 282 μg) (200 mM in dry DMSO, ProteoChem, c1102) were added to 1 eq (1 μmol, 660 μg) Biotin-(L-Asp)₃-β-Ala **1**. In the following, the reaction mixture was either added to 1.1 eq (1.1 μmol, 211 μg) 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide hydrochloride (EDC-HCl) (200 mM in dry DMSO, ProteoChem, c1100) for reaction in solution or to an excess of 3-10 eq N,N'-dicyclohexylcarbodiimide (DCC) resin (1.6 mmol/g, 150-300 μm, Agilent Technologies, PL3417-1689). To achieve effective swelling of the hydrophobic beads, the reaction solution was 2.5-fold diluted with chloroform (Roth, 7331). The reaction was carried out for 2.5 h at room temperature. To separate the reaction product from the DCC resin, the beads were pelletized at 1000×g for 30 s. After separating the beads from the supernatant, beads were washed twice by 5 min incubation with 10 μL DMSO; washes were combined with the supernatant from the reaction. The reaction products Biotin-(L-Asp)₃-β-Ala-NHS **2** or Biotin-(L-Asp)₃-β-Ala-sNHS **3** could be used for labelling of proteins without further purification. The NHS-activation of Biotin-(L-Asp)₃-β-Ala **1** was performed directly before use. To assess the storability of the reagent, the reaction mixture was stored for up to one week at -20°C after the incubation period of 2.5 h at room temperature. Aliquots of 1 μmol biotinylation reagent have been prepared for differential biotinylation of BSA or HeLa cells to assess the reactivity (see chapters 3.2, 3.3.2). In the following, the activated biotinylation reagents are named NHS-β-Ala-(L-Asp)₃-biotin **2** and Sulfo-NHS-β-Ala-(L-Asp)₃-biotin **3** according to literature.^[12]

NHS-β-Ala-(L-Asp)₃-biotin **2** (C₂₉H₃₉N₇O₁₅S₁): MW 757.73 g/mol.

Sulfo-NHS-β-Ala-(L-Asp)₃-biotin **3** (C₂₉H₃₈N₇NaO₁₈S₂): MW 859.76 g/mol.

3.1.1.3. Synthesis of Biotin-(L-Asp)₃-Cys-CONH₂

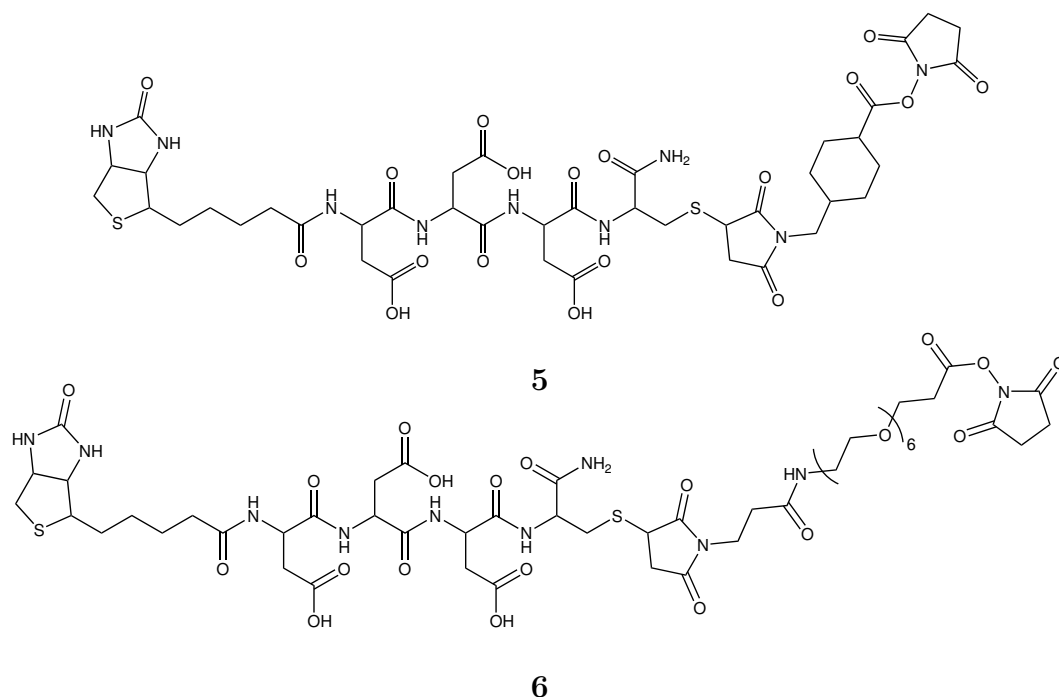


4

Biotin-(L-Asp)₃-Cys-CONH₂ **4** was kindly provided by Dr. Yixin Zhang (B Cube, Dresden). Peptide synthesis was performed by solid-phase synthesis on an Intavis peptide synthesizer using Fmoc chemistry based on a TentaGel S Ram resin (0.23 mmol/g). C-terminal activation of amino acid compounds was achieved with 0.5 M HBTU in DMF. N-terminal deprotection was performed in 20% piperidine in DMF. Cleavage from the resin and deprotection of aspartic acid side chains was achieved with a mixture of 90% TFA, 5% TIPS, 2.5% water and 2.5% dithiothreitol (DTT) for 2.5 h at room temperature. After precipitation in cold anhydrous diethyl ether, the peptide was dried under vacuum, followed by preparative reversed-phase HPLC (Luna[®] C18 column, 10 µm particle size, 30 x 250 mm, Phenomenex) with a 30 min linear water/acetonitrile gradient (0.1% TFA, 20 mL/min) for purification. For monitoring purposes, wavelengths of the UV/VIS diode array detector (Varian ProStar 325) were set to a range of 214-280 nm. Biotin-(L-Asp)₃-Cys-CONH₂ **4** was lyophilized and the purity was confirmed by an analytical Waters UPLC/MS system (Acquity BEH C18 column, 1.7 µm particle size, 2.1 x 50 mm, Waters) with TQ and TUV detector. Four synthesis batches of 25-270 mg peptide product were used for the experiments presented in this thesis.

Biotin-(L-Asp)₃-Cys-CONH₂ **4** (C₂₅H₃₇N₇O₁₂S₂): MW 691.73 g/mol.

3.1.1.4. Coupling of Biotin-(L-Asp)₃-Cys-CONH₂ to SMCC-Crosslinker or SM(PEG)₆-Crosslinker



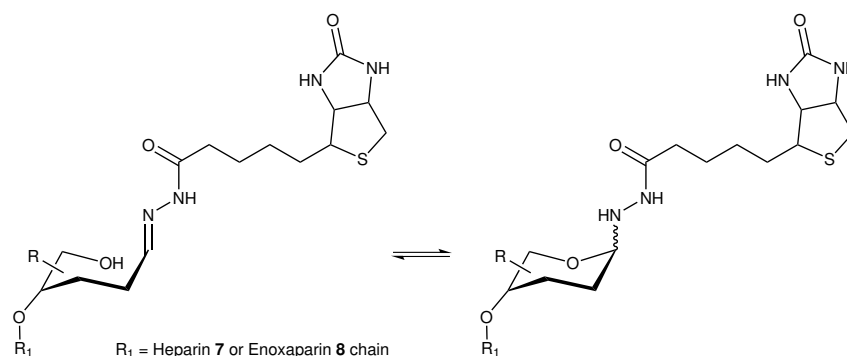
Biotin-(L-Asp)₃-Cys-CONH₂ **4** was dissolved in dry DMSO (over molecular sieve, Sigma-Aldrich, 41648) to a final concentration of 300 mg/mL (434 mM). The stock solution could be stored under inert gas at -20°C for several months. Before coupling to the crosslinker, 0.5 eq (0.5 μmol, 143 μg) tris(2-carboxyethyl)phosphine (TCEP) (200 mM in dry DMSO, BioVision, 1202) were incubated with 1 eq (1 μmol, 692 μg) peptide educt **4** for 10 min at room temperature. 1.3 eq (1.3 μmol, 435 μg) succinimidyl-4-(N-maleimidomethyl)cyclohexane-1-carboxylate (SMCC) (300 mM in dry DMSO, ProteoChem, c1108) or 1.3 eq (1.3 μmol, 782 μg) succinimidyl-([N-maleimidopropionamido]hexaethyleneglycol)ester (SM(PEG)₆) (300 mM in dry DMSO, Thermo Scientific, 22105) were added and the crosslinking reaction was carried out for 2.5 h at room temperature. The NHS-activation of Biotin-(L-Asp)₃-Cys-CONH₂ **4** via the SMCC- or SM(PEG)₆-crosslinker was always performed directly before use. The reaction products **5,6** could be used for labelling of proteins without further purification. 1 μmol (1.03 mg/1.29 mg) reagent was prepared for differential biotinylation of BSA for reactivity assessment (see chapter 3.2). For mass spectrometric validation experiments, 5 μmol (5.13 mg/6.47 mg) biotinylation reagent per cell culture flask were used for *in vitro* biotinylation of HeLa cells (see chapter 3.3.2) and 25 μmol (25.65 mg/32.33 mg) biotinylation reagent were prepared per mouse for *in vivo* perfusion (see chapter 3.4.2). In the following, the activated reagents are referred to as SMCC-Cys-(L-Asp)₃-biotin and SM(PEG)₆-Cys-(L-Asp)₃-biotin.

SMCC-Cys-(L-Asp)₃-biotin **5** (C₄₁H₅₅N₉O₁₈S₂): MW 1026.06 g/mol.

SM(PEG)₆-Cys-(L-Asp)₃-biotin **6** (C₅₁H₇₆N₁₀O₂₅S₂): MW 1293.33 g/mol.

3.1.2. Heparin-Based Biotinylation Reagents

3.1.2.1. Synthesis of Biotin-Heparin and Biotin-Enoxaparin

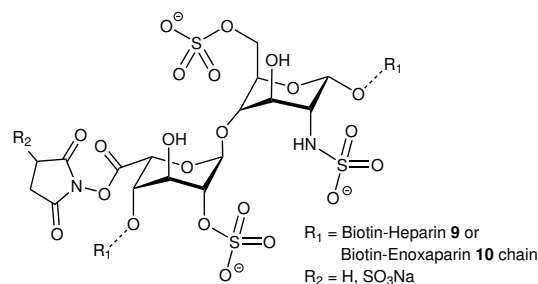


7,8

Heparin sodium salt from porcine intestinal mucosa (Calbiochem, 375095) is a highly sulfated polysaccharide with a mass range of 13 000-17 000 g/mol (average mass 15 000 g/mol). The most common disaccharide unit is composed of 2-O-sulfated iduronic acid and 6-O- and N-sulfated glucosamine (IdoA(2S)-GlcNS(6S)). Enoxaparin sodium (Dongying Tiandong Biochemical Industry Co., Dongying, China) is a low-molecular weight heparin with an average mass of 4421 g/mol (70% 2000-8000 g/mol, 20% <2000 g/mol) obtained by alkaline depolymerization of the benzyl ester derivative of heparin from porcine intestinal mucosa. The non-reducing end of the enoxaparin chain consists of a 4-enopyranose urinate structure, the degree of sulphatation is about 2 per disaccharide unit.

For biotinylation of the reducing end of the heparin chains, heparin sodium or enoxaparin sodium salt were dissolved in 0.1 M sodium acetate buffer, pH 5.0 (Sigma-Aldrich, S8750) to a final concentration of 25 mM or 50 mM, respectively. 10 eq (10 μ mol, 2.58 mg) biotin hydrazide (75 mM in dry DMSO, ProteoChem, b2106) were added to 1 eq (1 μ mol, 15 mg) heparin or 1 eq (1 μ mol, 4.4 mg) enoxaparin. The reaction solution was incubated with 100 eq (100 μ mol, 9.3 mg) aniline (Sigma-Aldrich, 242284) for 5 days at 70°C. The protocol was scaled to 5 μ mol or 25 μ mol Biotin-Heparin **7** or Biotin-Enoxaparin **8** if the reagents were used after consecutive activation for *in vitro* biotinylation of HeLa cells (see chapter 3.3.2) or *in vivo* perfusion per mouse (see chapter 3.4.2). The reaction products were purified by gel filtration. Biotin-Heparin **7** was purified using PD MiniTrap G-25 (GE Healthcare, 28-9180-07) or PD-10 desalting columns (GE Healthcare, 17-0851-01); Biotin-Enoxaparin **8** using PD MiniTrap G-10 (GE Healthcare, 28-9180-10) or PD MidiTrap G-10 (GE Healthcare, 28-9180-11) columns. For optimal recovery, the gravity protocols were used according to the manufacturer's instructions. For equilibration and elution, a five-fold dilution of PBS in water was used. Purified reaction products were dried in a vacuum concentrator (RVC 2-25 CDplus, Christ) and stored at -20°C.

3.1.2.2. NHS- or sNHS-Activation of Biotin-Heparin and Biotin-Enoxaparin



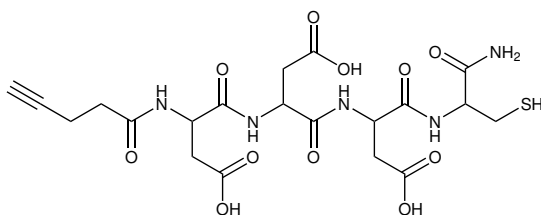
9,10

The heparin chain with an average mass of 15 000 g/mol (4421 g/mol for enoxaparin) consists on average of 25 (8 for enoxaparin) disaccharide units with one free carboxylic group each. For carboxylic group activation, Biotin-Heparin **7** or Biotin-Enoxaparin **8** were dissolved in PBS to a final concentration of 50 mM. 25 eq (25 μmol , 2.9 mg) NHS (500 mM in water, ProteoChem c1101) or 25 eq (25 μmol , 5.4 mg) sNHS (500 mM in water, ProteoChem c1102) and 25 eq (25 μmol , 4.8 mg) EDC-HCl (500 mM in water, ProteoChem c1100) were incubated with 1 eq (1 μmol , 15.2 mg) Biotin-Heparin **7** for 2.5 h at room temperature. 1 eq (1 μmol , 4.7 mg) Biotin-Enoxaparin **8** was incubated with 10 eq (10 μmol , 1.2 mg) NHS (500 mM in water, ProteoChem c1101) or 10 eq (10 μmol , 2.2 mg) sNHS (500 mM in water, ProteoChem c1102) and 10 eq (10 μmol , 1.9 mg) EDC-HCl (500 mM in water, ProteoChem c1100) for 2.5 h at room temperature. The reaction products **9,10** could be used for labelling of proteins without further purification. For mass spectrometric validation experiments, 5 μmol biotinylation reagent per cell culture flask were synthesized for the *in vitro* biotinylation of HeLa cells (see chapter 3.3.2) and 25 μmol biotinylation reagent were prepared per mouse for *in vivo* perfusion (see chapter 3.4.2). For *in vitro* and *in vivo* experiments, the activation was always performed with sNHS. In the following, those activated reagents are referred to as Biotin-Heparin-sNHS **9** and Biotin-Enoxaparin-sNHS **10**.

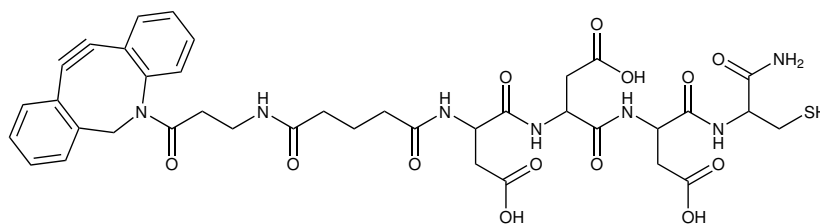
Activity tests with NHS- or sNHS-activated heparin compounds were performed using Biotin-Heparin **7** (Biotin-Enoxaparin **8**) or unbiotinylated heparin salt (enoxaparin salt) as educt. Differential ratios of NHS- or sNHS excess were tested using the activation protocol described above.

3.1.3. Click Chemistry-Based Reagents

3.1.3.1. Synthesis of Alkyne-(L-Asp)₃-Cys-CONH₂



11



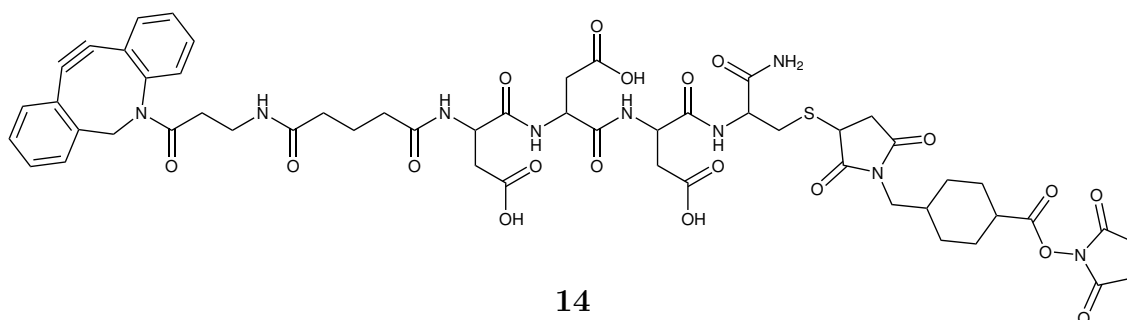
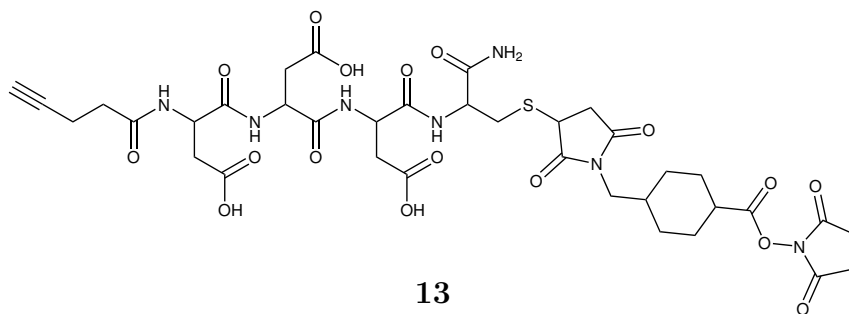
12

Alkyne-(L-Asp)₃-Cys-CONH₂ **11** and **12** were kindly provided by Dr. Yixin Zhang (B Cube, Dresden). Peptide synthesis was performed by solid-phase synthesis on an Intavis peptide synthesizer using Fmoc chemistry based on a TentaGel S Ram resin (0.23 mmol/g). C-terminal activation of amino acid compounds was achieved with 0.5 M HBTU in DMF. N-terminal deprotection was performed in 20% piperidine in DMF. Alkyne compounds used were 4-pentynoic acid (PA) and dibenzylcyclooctyne (DBCO) acid. Cleavage from the resin and deprotection of aspartic acid side chains was achieved with a mixture of 90% TFA, 5% TIPS, 2.5% water and 2.5% DTT for 2.5 h at room temperature. After precipitation in cold anhydrous diethyl ether, the peptide was dried under vacuum, followed by preparative reversed-phase HPLC (Luna[®] C18 column, 10 µm particle size, 30 x 250 mm, Phenomenex) with a 30 min linear water/acetonitrile gradient (0.1% TFA, 20 mL/min) for purification. For monitoring purposes, wavelengths of the UV/VIS diode array detector (Varian ProStar 325) were set to a range of 214-280 nm. PA-Alkyne-(L-Asp)₃-Cys-CONH₂ **11** (DBCO-Alkyne-(L-Asp)₃-Cys-CONH₂ **12**) were lyophilized and the purity was confirmed by an analytical Waters UPLC/MS system (Acquity BEH C18 column, 1.7 µm particle size, 2.1 x 50 mm, Waters) with TQ and TUV detector. One synthesis batch of 18 mg (11 mg) peptide product was used for the experiments presented in this thesis.

PA-Alkyne-(L-Asp)₃-Cys-CONH₂ **11** (C₂₀H₂₇N₅O₁₁S): MW 545.52 g/mol.

DBCO-Alkyne-(L-Asp)₃-Cys-CONH₂ **12** (C₃₈H₄₃N₇O₁₃S): MW 837.86 g/mol.

3.1.3.2. Coupling of Alkyne-(L-Asp)₃-Cys-CONH₂ to SMCC-Crosslinker



PA-Alkyne-(L-Asp)₃-Cys-CONH₂ **11** or DBCO-Alkyne-(L-Asp)₃-Cys-CONH₂ **12** was dissolved in dry DMSO (over molecular sieve, Sigma-Aldrich, 41648) to a final concentration of 300 mM. Before coupling to the crosslinker, 0.5 eq (0.5 μmol, 143 μg) TCEP (50 mM in dry DMSO, BioVision, 1202) were incubated with 1 eq (1 μmol, 545 μg or 838 μg) peptide educt **11** or **12** for 10 min at room temperature. 1.3 eq (1.3 μmol, 435 μg) SMCC (300 mM in dry DMSO, ProteoChem, c1108) were added and the crosslinking reaction was carried out for 2.5 h at room temperature. The activation via the SMCC-crosslinker was always performed directly before use. The reaction products **13,14** could be used for labelling of proteins without further purification. 0.5 μmol (440 μg/586 μg) reagent was prepared for differential biotinylation of BSA for reactivity assessment (see chapter 3.2). In the following, the activated reagents are referred to as SMCC-Cys-(L-Asp)₃-DBCO-alkyne and SMCC-Cys-(L-Asp)₃-PA-alkyne.

SMCC-Cys-(L-Asp)₃-PA-alkyne **13** (C₃₆H₄₅N₇O₁₇S): MW 879.85 g/mol.

SMCC-Cys-(L-Asp)₃-DBCO-alkyne **14** (C₅₄H₆₁N₉O₁₉S): MW 1172.19 g/mol.

3.1.4. Commercial Biotinylation Reagents

All validation experiments have been performed in comparison to the commercial reagents Sulfo-NHS-LC-biotin (ProteoChem, b2103) and NHS-PEG₁₂-biotin (Thermo Scientific, 21313). Sulfo-NHS-LC-biotin was stored in solid form at +4°C, NHS-PEG₁₂-biotin stock solution was prepared as 250 mM stock in dry DMSO (over molecular sieve, Sigma-Aldrich, 41648) and stored at -20°C for several months. Dilutions in PBS or water were made directly before use, concentrations were adapted to those of the synthesized reagents. For mass spectrometric validation experiments, 5 µmol (2.78 mg/4.71 mg) biotinylation reagent Sulfo-NHS-LC-biotin or NHS-PEG₁₂-biotin per cell culture flask were used for *in vitro* biotinylation of HeLa cells (see chapter 3.3.2) and 25 µmol (13.91 mg/23.53 mg) for *in vivo* perfusion per mouse (see chapter 3.4.2).

Sulfo-NHS-LC-biotin (C₂₀H₂₉N₄NaO₉S₂): MW 556.58 g/mol.

NHS-PEG₁₂-biotin (C₄₁H₇₂N₄O₁₈S): MW 941.10 g/mol.

3.2. Reactivity Assessment

3.2.1. Reaction with BSA as Model Protein

1 mg/ml bovine serum albumin (BSA) stock (Sigma, A2153) was prepared in phosphate buffered saline (PBS) (Gibco, 18912-014). Synthesized or commercial biotinylation or click chemistry based reagents (see chapter 3.1) were dissolved or diluted in water directly before use to a final concentration of 10 mM. The reagents were added to 100 µg BSA aliquots (1.505 nmol, 1 eq) in different ratios (1 eq - 100 eq). The reaction was quenched after 15 min incubation at room temperature using ten-fold excess (10 eq - 1000 eq) of 2-amino-2-hydroxymethyl-propane-1,3-diol (Tris) solution (100 mM, in water, Sigma-Aldrich, T1503). 10% TFA (in water, ProteoChem, LC6203) was added to a final concentration of 0.1% TFA. Biotinylated BSA samples were desalted using 100 µL OMIX C4 pipette tips (Agilent Technologies, A57009100). The desalting tips were wetted twice with 100 µL 50% UPLC/MS grade acetonitrile (Biosolve, 012041) in UPLC/MS grade water (Biosolve, 232141), followed by equilibration with two times 100 µL 0.1% TFA in UPLC/MS grade water. Proteins were bound to the resin by aspirating the sample solution twenty times into the tip. The desalting tips were washed three times with 100 µL 0.1% TFA in water. Elution was performed by aspirating ten times 100 µL 75% acetonitrile, 0.1% TFA in water. Samples were dried in a vacuum concentrator (RVC 2-25 CDplus, Christ), redissolved by 5 min sonication in 10 µL 5% acetonitrile, 0.1% TFA in water and co-crystallized with a two- to four-fold excess of sinapinic acid (10 mg/mL, in 50% acetonitrile, 0.1% TFA in water, ProteoChem, p9102) onto an Opti-TOF™ 384 matrix-assisted laser desorption/ionization (MALDI) target plate (AB SCIEX). At least 8 spots per sample were prepared.

In addition to the testing of different reagents in various ratios, other parameters of the biotinylation reaction have been examined using the protocol described. To assess the storability of the reagents, the reaction mixture in dry DMSO was stored for up to one week at -20°C before coupling to the model protein BSA. Moreover, different incubation times before quenching (30 s - 1 h) as well as different reaction buffers (PBS buffer, pH 7.45 (Gibco, 18912-014) / MES buffer, pH 6.0 (100 mM, Sigma-Aldrich, M8250) / Sodium carbonate buffer, pH 9.2 (100 mM, Sigma-Aldrich, S7795)) were examined.

3.2.2. MALDI MS Linear High Mass Positive Mode

Mass spectrometric analysis of BSA samples was performed on a MALDI TOF/TOF™ 5800 system (AB SCIEX) using the TOF/TOF™ Series Explorer™ Software (AB SCIEX, version V4.1.0) with oracle database schema version 4.0.0, data version 4.0.5. Measurements were performed in linear mode with MS acquisition method *Linear High Mass Positive*. Operating mode was set to *MS Linear High Mass Positive*, the mass range to 10 000-140 000 Da (focus mass 66 000 Da). 250 shots per sub-spectrum were acquired, 4 subspectra were accumulated (total: 1000 shots per spectrum). During acquisition, the sample plate was moved with continuous stage motion with a stage velocity of 600 µm/s. The digitizer used 20 ns bin size with a vertical scale of 0.05 (V full scale), a vertical offset of -2.5 (% full scale) and an input bandwidth of 20 MHz. Laser pulse rate was set to 400 Hz. Detector voltage multiplier was always adjusted following general maintenance of the instrument. The Y2 and X2 deflector voltages and the laser intensity were manually adjusted before sample measurement using the MS processing method *Linear High Mass Internal* with the m/z of BSA (66 431.000) as reference mass. 100 m/z were taken as mass tolerance, maximal outlier error was 5 m/z. For sample measurements, the MS processing method *Linear High Mass Default* defined the following peak detection characteristics: Minimal signal-to-noise ratio (S/N) was set to 10, the local noise window width to 250 m/z and the minimal peak width at the FWHM to 2.9 bins. The default settings were used for calibration with BSA as reference mass on calibration spots.

3.2.3. Analysis of MALDI MS Linear Mode Data

MS spectra (.t2d files) were analysed using the Data Explorer® Software (Applied Biosystems/MDS Analytical Technologies, Version 4.10 build 124). The Noise Filter/Smooth function was used for Gaussian peak smoothing. m/z at the peak maximum of the singly charged peak was extracted, average and the standard deviation was calculated over the spectra acquired from the same sample. The difference to the m/z of the unmodified BSA and the mass of the added reagent tag were used to calculate the number of reagent tags added to one BSA molecule.

3.3. *In Vitro* Validation

3.3.1. Cell Culture

HeLa cell line (luc+/GFP+) was obtained from Dr. Thomas Höfner (HI-STEM gGmbH, Heidelberg, Germany). Cell line authenticity was confirmed by Multiplexion GmbH (Heidelberg, Germany) via SNP-profiling of extracted DNA (DNeasy Blood & Tissue Kit, Qiagen, 69504). The adherent cells were cultured in T75 yellow cap flasks (TPP, 90076) at 37°C and 5% CO₂ (Heracell™ 240i CO₂ incubator, Thermo Scientific) in Iscove's Modified Dulbecco's Medium (IMDM) medium (Gibco, 21980-032) supplemented with 10% fetal calf serum (FCS) (PAN Biotech, 3302-P273105). After reaching 70-80% confluence, cells were washed with 10 mL PBS (Sigma, D8537) and detached with 5 mL 0.05% Trypsin/EDTA (Gibco, 25300-054) for 5 min at 37°C. The detaching solution was diluted with 5 mL of the cell culture medium, the cell suspension was transferred to a 15 mL Falcon tube (TPP, 91015) and cells were pelletized at 300×g for 5 min. Cells were resuspended in 10 mL cell culture medium and seeded in fresh cell culture flasks with 9 mL pre-warmed cell culture medium in a ratio of 1 to 10.

For cryopreservation, HeLa cells were washed and detached as described previously. Pelletized cells were resuspended in cell culture medium supplemented with 5% DMSO (Sigma-Aldrich, 276855) and aliquoted in Cryo vials (Thermo Scientific, 375353). Cells suspensions were cooled to -80°C within one day in a Mr. Frosty™ Freezing Container (Thermo Scientific, 5100) and long-term stored in liquid nitrogen. For thawing and cell recovery, cells were pre-warmed to 37°C and immediately transferred to 10 mL cell culture medium. Cells were pelletized at 300×g for 5 min, resuspended in 10 mL cell culture medium and seeded in a cell culture flask.

3.3.2. *In Vitro* Biotinylation

24 h prior to *in vitro* biotinylation on the cell culture flask for subsequent proteomic analysis, HeLa cells were washed, detached and pelletized as described in chapter 3.3.1. Cells were resuspended in cell culture medium and counted with a Neubauer chamber (Brand) using a two-fold dilution with 0.4% Trypan Blue Solution (Gibco, 15250). About 2.5×10^6 viable cells were seeded per cell culture flask.

For cell counting and the preparation of a negative control, four cell culture flasks were washed twice with 10 mL PBS. Cells were detached with 10 mL 10 mM ethylenediaminetetraacetic acid (EDTA) (Sigma, 03609) in PBS (Gibco, 18912014) within 10 min at 37°C. After pelletizing at 300×g for 5 min, HeLa cells were resuspended in PBS. The cell number was determined with a Neubauer chamber to about 5×10^6 viable cells per cell culture flask.

For *in vitro* biotinylation on the cell culture flask, HeLa cells were washed twice with PBS (Sigma, D8537) and overlaid with 4 mL PBS. Biotinylation reagents were dissolved or diluted with 1 mL PBS directly before use and applied to the adherent HeLa cells in a final concentration of 5 µmol per cell culture flask. Cells were

incubated with the reagents for 5 min at 37°C before quenching the reaction with a ten-fold excess of a 0.1 M Tris solution (Sigma, T1503) in water. Subsequently, the adherent cells were washed twice with 10 mL PBS followed by detachment with 10 mL 10 mM EDTA (Sigma, 03609) in PBS (Gibco, 18912014) within 20 min at 37°C. After pelletizing at 300×g for 5 min, HeLa cells were resuspended in PBS.

During optimization of the biotinylation procedure with Biotin-Heparin-sNHS **9**, 5 min pre-incubation of HeLa cells with 2 mg salmon sperm DNA (Invitrogen, 15632-011) in PBS before biotinylation was tested.

For assessment of the influence of the biotinylation procedure on the cell viability, living and dead cells were counted in PBS after the detachment with a Vi-cell™ XR (Beckmann Coulter) based on the Trypan Blue method. The biotinylation of adherent cells on cell culture flask was performed with two commercial reagents Sulfo-NHS-LC-biotin (ProteoChem, b2103) and NHS-PEG₁₂-biotin (Thermo Scientific, 21313), two peptide-based reagents SMCC-Cys-(L-Asp)₃-biotin **5** and SM(PEG)₆-Cys-(L-Asp)₃-biotin **6** and two heparin-based reagents Biotin-Heparin-sNHS **9** and Biotin-Enoxaparin-sNHS **10**. Negative control cells were only treated with pure PBS.

Whereas the biotinylation of cells for proteomic analysis was performed on adherent cells as described previously, cells for subsequent FACS analysis for reactivity assessment by labelling with different amounts of biotinylation reagent were biotinylated in suspension to ensure exact and equal cell count per experiment. After washing twice with PBS on the cell culture flask, HeLa cells were detached with 10 mL 10 mM EDTA (Sigma, 03609) in PBS (Gibco, 18912014) within 10 min at 37°C. After pelletizing at 300×g for 5 min, cells were resuspended in PBS, counted with a Neubauer chamber and equally aliquoted. Biotinylation reagents were dissolved or diluted with PBS to a final concentration of 10 mM. 1×10⁶ cells in 100 µL PBS were incubated with different amounts (ratios) of biotinylation reagents: 0.06 nmol (1 to 0.01), 0.6 nmol (1 to 0.1), 6 nmol (1 to 1), 60 nmol (1 to 10)). Ratios refer to a median protein amount of 300 µg (6 nmol) protein content per 1×10⁶ cells with a median protein weight of 50 kDa.^[130] The reaction was quenched with a ten-fold excess of a 0.1 M Tris solution in water. Cells were pelletized, washed twice with PBS and resuspended in FACS buffer (2% FCS (PAN Biotech, 3302-P273105) in PBS).

3.3.3. Visualisation of Biotinylation

3.3.3.1. FACS Analysis

As a control of the biotinylation of adherent HeLa cells, fluorescence-activated cell sorting (FACS) analysis was performed after biotinylation, detachment and cell counting in PBS (see chapter 3.3.2). 500 000 cells were pelletized at 300×g for 5 min at 4°C, resuspended in 500 µL FACS buffer (2% FCS in PBS) and kept on ice. Staining solution was prepared under light protection as fifty-fold dilution of Streptavidin Pacific Blue™ conjugate (Invitrogen, S-11222) in FACS buffer. For staining,

cells were pelletized and incubated with 10 μ L staining solution per 500 000 cells for 30 min on ice in the dark. After washing with 1 mL FACS buffer, cells were resuspended in 250 μ L FACS buffer and filtered into FACS tubes (BD Falcon, 352053) for analysis.

Differential FACS analysis of different amounts of biotinylation reagents on HeLa cells (see chapter 3.3.2) was also performed by staining the biotinylated cells with Streptavidin Pacific Blue conjugate according to the protocol described.

Cells were analysed at a medium flow rate with a CyAnTM ADP flow cytometer (Beckmann Coulter) equipped with 488 nm, 405 nm and 635 nm lasers. At least 10 000 events were recorded using a forward scatter threshold of 0.1%. For Pacific Blue, data were collected with the 405 nm laser 450/50 bandpass filter. PMT voltage of the fluorescent channel was adapted with unstained cells to set negative events to the first decade (log-scale). Flow cytometry data were recorded in peak, area and log parameters with the CyAnTM ADP Software Summit 4.3 and exported in .fcs format. Further analysis was performed with FlowJo (version 9.8.5). Viable cells and singlet cells were gated based on forward scatter (FSC) and side scatter (SSC) before histogram analysis of Pacific Blue staining.

3.3.3.2. Confocal Laser Scanning Microscopy

For visualization of plasma membrane protein biotinylation, 1×10^6 HeLa cells in cell culture medium were seeded in a cell culture insert (1.0 μ m, Falcon, 353102) within a MultiwellTM 6-well plate (Falcon, 353502) and grown at 37°C and 5% CO₂ as described in chapter 3.3.1. After 24 h, cell culture medium was removed and cells were washed twice with PBS. 3 μ mol biotinylation reagent (Sulfo-NHS-LC-biotin, NHS-PEG₁₂-biotin, SMCC-Cys-(L-Asp)₃-biotin, SM(PEG)₆-Cys-(L-Asp)₃-biotin) were diluted in 1.5 mL PBS directly before use, pure PBS served as negative control. Cell inserts were placed on 100 μ L of the biotinylation solution on a clean surface, cells were covered with the rest of the solution and incubated for 5 min at room temperature. Cells and cell culture inserts were washed three times with 2 mL PBS, followed by 10 min fixation with 1 mL ice-cold methanol (Sigma-Aldrich, 34860) at -20°C. After washing the cells and the cell insert with PBS, the cell insert membrane was cut into four parts and stored overnight in 0.1 M glycine solution (Sigma, G8898) in PBS in a 24-well cell culture plate (Greiner Bio-One, 662160) at 4°C.

To perform staining for confocal laser scanning microscopy, membrane pieces were washed for 15 min at room temperature in 0.2% TritonTM X-100 (Sigma-Aldrich, X100) followed by 5 min in PBS. After transfer to parafilm in a moist chamber, blocking was performed with 100 μ L 10% goat serum (Sigma, G6767) in PBS for 30 min at room temperature. Staining solution was prepared as 200-fold dilution of 4',6-diamidino-2-phenylindole (DAPI) (5 mg/ml in water, Sigma, D9542) and 200-fold dilution of Streptavidin Alexa Fluor 488 conjugate (life technologies, S32354) in 3% BSA in PBS. Membranes were washed with PBS and incubated with staining solution in a moist chamber for 1 h at room temperature. Subsequent to three

washing steps in 2 mL PBS, membrane pieces were transferred to a glass slide (SuperFrost® Plus, Thermo, J1800AMNZ), mounted with Fluorescent Mounting Medium (Dako, S3023), covered with a glass lid and stored under light protection at 4°C.

Slides were analysed with a LSM 700 Confocal Laser Scanning Microscope (Zeiss) operated with ZEN 2012 software (black edition, Zeiss, version 8.0.5.273) using the 405 nm and 488 nm laser diodes and the EC-Plan-Neofluar 40x/1.3 oil DIC objective (Zeiss, 420462-9900). Pictures were exported in .lsm format, scale bars were introduced using the ZEN lite 2011 software (blue edition, Zeiss, version 6.1.7601). The pictures presented in this thesis for direct comparison were taken using the same laser characteristics and the same optimized contrast properties in the individual channels for all samples within the analysis.

3.3.4. Cell Lysis

After biotinylation, washing and detachment (see chapter 3.3.2), biotinylated HeLa cells were pelletized at 300×g for 5 min and resuspended in radioimmunoprecipitation assay (RIPA) buffer (see Table 3.1) containing cOmplete EDTA-free protease inhibitor cocktail (Roche, 04693132001) to a final concentration of 1×10⁶ cells per 500 µL. The suspension was incubated on ice for 2 h and slightly vortexed from time to time. Samples were stored at -20°C overnight. Lysates were thawed and homogenized on ice 3 times 30 s with a T 10 basic ULTRA-TURRAX® homogenizer (IKA) at 20 000 rpm. Afterwards, samples were sonicated for 2 min with 1 s pulses at 30% amplitude with the Sonifier® W-250D (Branson). Cell lysates were centrifuged for 20 min at 20 000×g and 4°C, supernatants were kept at -20°C for downstream analysis (see chapter 3.6.1).

Concentration	Reagent	Company	Ordering No.
50 mM	Tris	Sigma-Aldrich	T1503
150 mM	NaCl	Sigma-Aldrich	31434
1%	NP-40	Sigma-Aldrich	74385
0.5%	Sodium deoxycholate	Sigma-Aldrich	30970
0.1%	SDS	Sigma-Aldrich	L6026
	water (deionized)		
1 tablet/50 ml	Protease inhibitor cocktail	Roche	04693132001

Table 3.1.: RIPA buffer for cell lysis.

3.4. *In Vivo* Validation

3.4.1. Mouse Strain

NOD.Cg-*Prkdc*^{scid} *Il2rg*^{tm1Wjl}/SzJ (NSG) mice from The Jackson Laboratory (Bar Harbor, USA) were bred in the German Cancer Research Center (DKFZ) animal facility. Mice were kept under specific pathogen-free conditions in individually ventilated cages. Experiments have been performed with untreated female mice at 18-26 weeks of age. All animal handling and procedures followed German legal regulations and were previously approved by the national authorities (Regierungspräsidium Karlsruhe, TVA number G134/12).

3.4.2. *In Vivo* Perfusion of Mice

The *in vivo* perfusion of the vascular system was performed according to the protocol of Rösli *et al.*^[185] Perfusion solutions were kept at 40°C, biotinylation reagents were dissolved directly before use. Mice were anaesthetized with a mixture of 200 µg Ketamin (Ketavet, Parke-Davis), 20 µg Xylazin (Rompun, Bayer) and 3 µg Acepromazin (Vetranquil, Ceva) per g of body weight. After complete loss of toe pinch reflexes, the animal was fixed with tape on a pre-warmed surgery table (Hugo Sachs Elektronik) with the head on the lower side. Abdominal cavity was opened by dissecting the skin and the peritoneum. A median sternotomy was performed followed by the insertion of a butterfly needle with a small barb into the left ventricle. The needle was connected to a syringe pump (LA-30, Landgraf Laborsysteme) via Heidelberger extensions (B. Braun, 4097300) and a three way valve (Discofix®, B. Braun, 4098102). A small cut was made in the right atrium to allow blood and perfusion solutions to flow out. The pump was set to a constant flow rate of 1.1 mL/min. Tubing was pre-filled with 2.5 mL filtered and pre-warmed perfusion solution (10% Dextran-40 000 MW (US Biological, D6030) in PBS). Perfusion was performed using 25 µmol biotinylation reagent in 11 mL pre-warmed perfusion solution, followed by perfusion with 11 mL of quenching solution (50 mM Tris (Sigma, T1503) in 10% Dextran in PBS). To inactivate non reacted biotinylation reagent after the whole body perfusion, the thoracic and abdominal cavity were continuously washed with sprinkling solution (50 mM Tris in PBS) during the perfusion with biotinylation reagent and during the first 5 min of the perfusion with the quenching reagent.

Mice were divided into eight perfusion groups: Perfusions were performed with two commercial reagents Sulfo-NHS-LC-biotin (ProteoChem, b2103) and NHS-PEG₁₂-biotin (Thermo Scientific, 21313), two peptide-based reagents SMCC-Cys-(L-Asp)₃-biotin **5** and SM(PEG)₆-Cys-(L-Asp)₃-biotin **6** as well as two heparin-based reagents Biotin-Heparin-sNHS **9** and Biotin-Enoxaparin-sNHS **10**. Negative control mice were either perfused with pure perfusion solution or stayed unperfused.

After perfusion, liver, kidneys, spleen, tongue and brain were excised, separated from connective tissue or fat and the renal capsule was removed. Small parts of the organs were embedded in Tissue-Tek® O.C.T.™ medium (Sakura, 4583) and

stored in embedding molds (Polysciences, 18986) at -80°C for subsequent immuno-histochemical analysis (see chapter 3.4.3.1). The main part of liver and kidney tissue was snap-frozen in liquid nitrogen and stored at -80°C for protein extraction and proteomic analysis (see chapters 3.4.4, 3.6.2).

3.4.3. Visualization of Biotinylation

3.4.3.1. Confocal Laser Scanning Microscopy

Small parts from kidney, liver, spleen, tongue and brain tissue of perfused mice (see chapter 3.4.2) embedded in Tissue-Tek® O.C.T.™ medium (Sakura, 4583) were cut into 8 µm sections using a Microm™ HM 525 cryotome (Thermo Scientific). Two sections per object were mounted on a glass slide (SuperFrost® Plus, Thermo, J1800AMNZ). Slide were stored at -80°C until staining.

Before staining, sections were fixed in acetone (Sigma-Aldrich, 32201) at -20°C for 8 min. After drying at room temperature, Daco silico pen (Dako, S2002) was applied around the individual sections. Samples were transferred to a moist chamber and blocked with 20% goat serum (Sigma, G6767) in PBS for one hour at room temperature. Slides were washed for 5 min with PBS and incubated at room temperature with a 200-fold dilution of primary antibody (Rat anti-mouse CD31 antibody, BD Biosciences, 550274) in 12% BSA (Sigma-Aldrich, A2153) in PBS in a moist chamber for 2 h. Control stains were treated on the same slide with 12% BSA in PBS without the primary antibody. The slides were rinsed three times with PBS before incubation with a 250-fold dilution of secondary antibody (Alexa Fluor 594 Goat Anti-Rat IgG, life technologies, A11007), 200-fold dilution of DAPI (5 mg/ml in water, Sigma, D9542) and 250-fold dilution of Streptavidin Alexa Fluor 488 conjugate (life technologies, S32354) in 12% BSA in PBS in a moist chamber for 1 h. The control stains were performed without Streptavidin Alexa Fluor 488 conjugate. After washing the slides three times with PBS, sections were mounted with Fluorescent Mounting Medium (Dako, S3023) and stored at 4°C under light protection.

Slides were analysed with a LSM 700 Confocal Laser Scanning Microscope (Zeiss) operated with ZEN 2012 software (black edition, Zeiss, version 8.0.5.273) using 405 nm, 488 nm and 555 nm laser diodes and the Plan-Apochromat 20x/0.8 objective (Zeiss, 420650-9901). Pictures were exported in .lsm format, scale bars were introduced using the ZEN lite 2011 software (blue edition, Zeiss, version 6.1.7601). The pictures presented in this thesis for direct comparison were taken using the same laser characteristics and the same optimized contrast properties in the individual channels for all samples within the analysis.

3.4.4. Protein Extraction of Kidney and Liver Tissue

Snap-frozen kidney and liver of perfused mice (see chapter 3.4.2) were weighed and hackled tissue was transferred into 15 ml Falcon tubes (TPP, 91015). Protein extraction buffer (see Table 3.2) was added to a final concentration of 25 µL per mg

kidney tissue and 20 μ L per mg liver tissue. Samples were homogenized for 10 min with a T 10 standard ULTRA-TURRAX[®] homogenizer (IKA) at 26 000 rpm followed by 2 min sonication with 1 s pulse time and 35% intensity using Sonifier[®] W-250D (Branson). Tissue homogenates were transferred to 2 ml tubes (eppendorf, 0030120094) and heated to 95°C for 20 min in a thermomixer (eppendorf). Remaining tissue debris was pelletized at 20 000 \times g for 20 min at room temperature. The supernatant was aliquoted and stored at -20°C for a maximum period of one year for proteomic analysis (see chapter 3.6.2).

Concentration	Reagent	Company	Ordering No.
2%	SDS	Sigma-Aldrich	L6026
50 mM	Tris	Sigma-Aldrich	T1503
10 mM	EDTA	Sigma-Aldrich	03609
	water (deionized)		
50 mM	HCl (ad pH 7.0)	Sigma-Aldrich	30721
1 tablet/50 ml	Protease inhibitor cocktail	Roche	04693132001

Table 3.2.: Protein extraction buffer for tissue samples.

3.5. Protein Concentration Determination of Lysates and Assessment of Biotinylation

3.5.1. Protein Concentration Determination

Protein concentrations of cell lysates and tissue homogenates were determined using the Pierce[™] BCA Protein Assay Kit (Thermo Scientific, 23225). A BSA standard curve (Thermo Scientific, 23209) was prepared in water in a concentration range from 1-2000 μ g/mL. Microplate procedure with 20 μ L sample in duplicate measurements was applied, working reagent was prepared as fifty-fold dilution of reagent B in reagent A according to the manufacturer's protocol. The microplate was incubated at room temperature for 30 min before measuring 562 nm absorbance with a plate reader (SpectraMax[®] M5, Molecular Devices) operated with SoftMax[®] Pro software (version 5.4). Cell lysates were measured undiluted, tissue homogenates in a five-fold dilution in water. Averaged sample values out of duplicate or triplicate measurements were corrected by sample buffer absorbance in the dilution used for the corresponding samples.

3.5.2. ELISA

The ELISA was performed on 96-well MaxiSorp Nunc-Immuno Polystyrene Plates (Thermo Scientific, 442404) or 96-well Half Area High Bind Polystyrene Plates

(Corning, 3690). 0.1 nmol protein in 100 μ L water were applied per well, i.e. 7.1 μ g kidney or liver protein according to an average protein mass of 71 000 Da.^[280]

The ELISA standard curve used for relative quantification of the biotin content among the samples was generated by combination of different ratios of biotinylated and unbiotinylated BSA (in water, 1 mg/mL stock each) to a total concentration of 0.1 nmol in 100 μ L (6.65 μ g in 100 μ L). For preparation of the ELISA standard, biotinylation of BSA in PBS (Sigma, A2153) was performed with Sulfo-NHS-LC-biotin in a 1 to 10 ratio as described in chapter 3.2.1. Biotinylated BSA was purified using Vivaspin 15R columns (Sartorius, VS15RH01) with three washing steps (15 mL water, 15 min, 3000 \times g) and three elution/rinsing steps using 5 mL water each. Protein concentration of the purified product was determined as described in chapter 3.5.1 and adapted with water to a final concentration of 1 mg/mL. Standard curve was prepared in a concentration range of 0-0.5 μ g biotinylated BSA with six narrow intervals from 0-0.1 μ g.

ELISA plates were coated with 0.1 nmol protein (ELISA standard curve or sample) in 100 μ L water per well and incubated overnight at 4°C. Wells were washed three times with PBS and free binding sites were blocked with 2% BSA in PBS (1 h, room temperature). After washing with PBS, 100 μ L Streptavidin-horseradish peroxidase (HRP) solution (1:1000 in 2% BSA in PBS, GE Healthcare, RPN 1231) was incubated on the wells for 1 h at room temperature. Wells were washed three times with 0.1% Tween[®] 20 (Sigma-Aldrich, P1379) in PBS and three times with PBS. For detection, 100 μ L (50 μ L on half area plates) BM Blue POD Substrate solution (Roche, 11484281001) were applied. The colour reaction was quenched after 5-20 min with 70 μ L (35 μ L on half area plates) 1 M sulfuric acid (Sigma-Aldrich, 258105). Absorbance at 450 nm subtracted by absorbance at 690 nm was measured with a plate reader (SpectraMax[®] M5, Molecular Devices) operated with SoftMax[®] Pro software (version 5.4). For analysis, averaged sample values out of duplicate or triplicate measurements were corrected by sample buffer absorbance.

3.5.3. SDS-PAGE and Western Blot

Sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) was conducted using the XCell SureLock[™] Mini-Cell Electrophoresis System (Invitrogen) with NuPAGE[®] Novex[®] 4-12% Bis-Tris, 1.5 mm, 10- or 15-well gels (Life Technologies) and 1x NuPAGE[®] MOPS SDS Running Buffer (Life Technologies, NP0001). 5x Loading Buffer (see Table 3.3) and, if performed under reducing conditions, 10x DTT (500 mM in water, Sigma-Aldrich, D9779) was added to the samples to a final 1x concentration and heated to 95°C for 10 min before loading on the gel. PageRuler[™] Plus Prestained Protein Ladder (Life Technologies, 26619) or Spectra[™] Multicolor Broad Range Protein Ladder (Life Technologies, 26634) were used as markers. Gels were run for 1 h at 180 V and stained with the one-step Coomassie based InstantBlue[™] gel stain (Expedeon, ISB1L) for 1 h at room temperature.

Concentration	Reagent	Company	Ordering No.
208 mM	Tris-HCl (pH 6.8)	Sigma-Aldrich	T1503, 30721
33%	Glycerine	Sigma-Aldrich	G6279
5%	SDS	Sigma-Aldrich	L6026
0.06%	Bromophenol Blue water (deionized)	Sigma-Aldrich	B8026

Table 3.3.: 5x Gel loading buffer.

Western Blots were performed as semi-wet transfer using the XCell II™ Blot Module (Invitrogen) with Immobilon-P PVDF transfer membranes (Millipore, IPVH-00010). Biotinylated protein samples as well as Biotin-Heparin-sNHS **9** were run in NuPAGE® Novex® 4-12% Bis-Tris, 1.5 mm, 10- or 15-well gels (Life Technologies) and blotted during 90 min at 30 V using 1x Transfer Buffer (see Table 3.4) containing 10% Methanol (Sigma-Aldrich, 34860). Free binding sites on the transfer membranes were blocked in 4% milk powder (Sigma-Aldrich, 70166) in PBS overnight at 4°C, followed by incubation with Streptavidin-HRP solution (1:1000 in 2% Milk Powder in PBS, GE Healthcare, RPN 1231) for 30 min at room temperature. Transfer membranes were washed twice with 0.1% Tween® 20 (Sigma-Aldrich, P1379) in PBS and twice with PBS. For detection, Amersham® ECL® Prime Western Blotting Detection Reagents A and B (GE Healthcare, RPN2232) were used in five-fold dilution in water. Western Blot CL-XPosure™ films (Thermo, 34090) were incubated on the transfer membrane in a film cassette (Hypercassette™, GE Healthcare, RPN12642) and developed in a Classic EOS film developing machine (AGFA).

Concentration	Reagent	Company	Ordering No.
0.5 M	Bicine	Sigma-Aldrich	B3876
0.5 M	Bis-Tris	Sigma-Aldrich	B9754
20 mM	EDTA	Sigma-Aldrich	03609
1 mM	Chlorobutanol hemihydrat water (deionized)	Sigma-Aldrich	112054

Table 3.4.: 20x Western blot transfer buffer.

3.6. Sample Preparation for Mass Spectrometric Analysis

3.6.1. Capturing of Surface Proteins out of Cell Lysates

Cell lysates of biotinylated HeLa cells were prepared and protein concentration was determined as described in chapters 3.3.4 and 3.5.1. Samples were processed by

capturing of biotinylated proteins out of 500 µg total protein amount on streptavidin-sepharose resin under reductive conditions, followed by alkylation, tryptic digest and desalting for proteomic analysis.

100 µL Streptavidin-sepharose slurry (GE Healthcare, 17-5113-01) per tube were prepared by washing three times with 500 µL 1% 4-nonylphenyl-polyethylene glycol (NP-40) (Sigma, 74385), 0.1% sodium dodecyl sulfate (SDS) (Sigma, L6026) in PBS (termed Wash Buffer in the following). Cell lysate with a total protein content of 500 µg was added to the washed streptavidin-sepharose. 2 mM TCEP (BioVision, 1202) in 20% SDS in PBS were added to a final concentration of 2% SDS, followed by 2 h tumbling (tube rotator, Cole-Parmer) at room temperature. After the incubation period, the streptavidin-sepharose with the captured proteins was separated from the supernatant by centrifugation at 300×g for 30 s at room temperature. After resuspending the streptavidin-sepharose in 500 µL Wash Buffer, the slurry was transferred to a centrifugal filter tube (Ultrafree[®]-MC SV Centrifugal Filters, Merck Millipore, UFC30SV00) for washing. The filter device was centrifuged at 300×g for 30 s, flow-through was discarded. The streptavidin-sepharose was washed twice with 150 µL 50 mM iodoacetamide (IAA) (Sigma-Aldrich, I1149) in PBS before resuspending in 150 µL of the IAA solution and transferring to a new tube. The filter tube was rinsed twice with another 150 µL of the IAA solution, yielding a total volume of about 500 µL streptavidin-sepharose in IAA solution. Samples were incubated for 30 min in the dark before quenching the alkylation reaction with 300 µL 100 mM L-cysteine (Sigma-Aldrich, 30089) in water. After 15 min of incubation, streptavidin-sepharose was separated from the supernatant by centrifugation at 300×g for 30 s, resuspended in 500 µL Wash Buffer and transferred to a centrifugal filter device. The streptavidin-sepharose was washed ten times with 500 µL 50 mM Tris-HCl (Sigma-Aldrich, T1503, 30721), 1 mM CaCl₂ (Sigma-Aldrich, C4901) in water, pH 8 (termed Tryptic Digestion Buffer in the following), before transferring it in two steps in a total of 1 ml Tryptic Digestion Buffer to a new tube. After centrifugation at 300×g for 30 s, supernatant was discarded and 200 µL Tryptic Digestion Buffer were added. The tryptic digest was carried out using 1.6 µg Sequencing Grade Modified Trypsin (20 µg vial in 250 µL Tryptic Digestion Buffer, Promega, V511A) for 10 h at 37°C at 1000 rpm. The modified enzyme provides resistance to autolytic digestion as well as high tryptic specificity as a consequence of tosylphenylalanylchloromethylketone (TPCK) treatment.

Following tryptic digest, the streptavidin-sepharose slurry was transferred to a centrifugal filter device and the digest was separated from the beads by centrifugation at 300×g for 1 min, the flow-through was kept. For desalting, 2.2 µL 10% TFA (ProteoChem, LC6203) were added to a final concentration of 0.1% TFA. Desalting was carried out using OMIX C18 pipette tips (100 µL, Agilent Technologies, A57003100). The desalting tips were wetted twice with 50% UPLC/MS grade acetonitrile (Biosolve, 012041) in UPLC/MS grade water (Biosolve, 232141), followed by two times equilibration with 0.1% TFA in UPLC/MS grade water. Proteins were bound to the resin by aspirating the sample solution twenty times into the tip. The desalting tips were washed three times with 0.1% TFA in water. Elution was per-

formed by aspirating ten times 75% acetonitrile, 0.1% TFA in water. Samples were dried in a vacuum concentrator (RVC 2-25 CDplus, Christ) and stored at -20°C for subsequent LC/MS analysis.

During optimization of the heparin sample capturing procedure, the following steps were included after alkylation of the captured proteins on the streptavidin-sepharose beads: (i) Washing with high-salt buffer: five washing steps with 0.1% SDS, 2 M NaCl (Sigma-Aldrich, 31434) in PBS; (ii) Washing with pH shift: 3× three washing steps, 10 mM phosphate buffer (Sigma-Aldrich, 71500, 30412, 222003), 0.1% SDS (pH 5, pH 8, pH 12); (iii) Carbonate extraction: tumbling for 1 h at 4°C in 1 mL 100 mM sodium carbonate (Sigma-Aldrich, S7795), pH 11; (iv) Delipidation: tumbling for 30 min at room temperature in 400 µL Wash Buffer, 800 µL 1-butanol/diisopropylether solution (40:60, Chemsolute, 2513, AppliChem, A4322).

3.6.2. Capturing of Vascular Accessible Proteins out of Kidney and Liver Tissue Homogenates

Homogenates of perfused mouse kidney and liver tissue were prepared and protein concentration was determined as described in chapters 3.4.4 and 3.5.1. Samples were processed by capturing of biotinylated proteins out of 1.25 mg total protein amount of kidney homogenate or 2.55 mg of liver homogenate on 100 µL streptavidin-sepharose resin under reductive conditions, followed by alkylation, delipidation, tryptic digest and desalting for proteomic analysis.

The procedure follows the protocol described in chapter 3.6.1, but includes a delipidation step: After alkylation and quenching with L-cysteine, 400 µL Wash Buffer and 800 µL 1-butanol/diisopropylether solution (40:60, Chemsolute, 2513, AppliChem, A4322) were added to the streptavidin-sepharose and tumbled for 30 min at room temperature. After centrifugation at 300×g for 30 s, the solvent phase and the water phase were discarded and the streptavidin-sepharose was washed once with 500 µL Wash Buffer. Following resuspension in 500 µL Wash Buffer, the streptavidin-sepharose slurry was transferred to a centrifugal filter device and washed with Tryptic Digestion Buffer as already described.

Desalting of kidney samples was carried out using OMIX C18 pipette tips (100 µL, Agilent Technologies, A57003100), whereas the liver samples were desalted in parallel using a 96-well, C18 Lab-in-a-plate Flow-Thru Plate (Glygen, FN18C18) with 3 min, 200×g centrifugation steps: Wells were wetted with 200 µL 50% UPLC/MS grade acetonitrile (Biosolve, 012041) in UPLC/MS grade water (Biosolve, 232141), followed by equilibration with 200 µL 0.1% TFA in UPLC/MS grade water. Samples were applied on the wells in three steps. The plate was washed three times with 0.1% TFA in water. Elution was performed with 100 µL 75% acetonitrile, 0.1% TFA in water in three steps. Samples were dried in a vacuum concentrator (RVC 2-25 CDplus, Christ) and stored in Protein LoBind tubes (Eppendorf, 022431081) at -20°C for subsequent LC/MS analysis.

During optimization of the procedure, the following steps were tested on kidney samples (capturing of biotinylated proteins out of 2.00 mg total protein): (i) Washing after alkylation with high-salt buffer: three steps, 1% NP-40 (Sigma, 74385), 0.1% SDS (Sigma, L6026), 2 M NaCl (Sigma-Aldrich, 31434) in PBS; (ii) Pre-blocking the streptavidin-sepharose before capturing: washing streptavidin-sepharose as described, 2 h tumbling at room temperature with 750 μ L 10 mg/mL BSA solution before capturing sample homogenate; (iii) Ultracentrifugation of samples before taking the supernatant for capturing: 1 h, 100 000 \times g (Sorvall Discovery M120SE, Thermo Scientific); (iv) Benzonase treatment of samples before capturing: chloroform/methanol precipitation (see chapter 3.6.3), resolubilisation in 4 M urea (Sigma-Aldrich, 33247), 0.1% TritonTM X-100 (Sigma-Aldrich, X100), 100 mM Tris (Sigma-Aldrich, T1503), 1 mM magnesium chloride (Sigma-Aldrich, M8266), incubation with 250 U benzonase nuclease (Novagen) per mL lysate for 30 min at room temperature.

3.6.3. Preparation of Full Proteome Samples from Kidney and Liver Tissue Homogenates

Protein concentration of kidney and liver homogenates of PBS-perfused or unperfused mice was determined as described in chapter 3.5.1. 200 μ g protein samples were processed by reduction, alkylation and precipitation, followed by tryptic digest and desalting for proteomic analysis.

Tissue homogenate with 200 μ g protein content was incubated in a 5 mM TCEP solution (50 mM in water, BioVision, 1202) for 30 min at 60°C. Alkylation was performed by incubation in 15 mM IAA solution (100 mM in water, Sigma-Aldrich, I1149) for 30 min at 37°C in the dark. 100 μ g protein aliquots per sample were precipitated with chloroform/methanol as described previously.^[281] Briefly, the protein solution was diluted with 4 volumes of methanol (Sigma-Aldrich, 34860), 1 volume chloroform (Roth, 7331.2), 3 volumes of water and centrifuged for 2 min at 15 000 \times g. The upper organic phase was discarded, 4 volumes of methanol were added and the proteins were pelletized for 2 min at 15 000 \times g. The protein pellets were resolubilized in 0.1% RapiGestTM SF Surfactant (Waters, 186001861) in Tryptic Digestion Buffer (50 mM Tris-HCl, 1 mM CaCl₂ in water, pH 8) by sonication (Ultrasonic Cleaner, VWR) for 20 min and digested with trypsin (1:50, w/w, Promega, V511A) for 16 h at 37°C and 1200 rpm. Samples were acidified to 0.5% TFA (10% TFA in water, ProteoChem, LC6203) and incubated for 30 min at 37°C at 1200 rpm. Acidified protein solutions were separated from insoluble detergent by-products by 10 min centrifugation at 20 000 \times g. Samples were desalted using a 96-well, C18 Lab-in-a-plate Flow-Thru Plate (Glygen, FNSC18) as described in chapter 3.6.2, dried in a vacuum concentrator (RVC 2-25 CDplus, Christ) and stored in Protein LoBind tubes (Eppendorf, 022431081) at -20°C for subsequent LC/MS analysis.

3.7. Mass Spectrometric Analysis

3.7.1. Discovery Experiment

3.7.1.1. UPLC-Based Separation

Prior to MALDI-MS analysis, samples were separated by a nanoACQUITY UPLC system (Waters) equipped with a nanocapillary, reversed-phase Peptide BEH C18 nanoACQUITY Column, 130 Å pore size, 1.7 µm particle size, 75 µm x 250 mm (Waters, 186003545) or a M-Class Peptide BEH C18 Column, 130 Å pore size, 1.7 µm particle size, 75 µm x 250 mm (Waters, 186007484, used for liver samples, see chapter 3.6.2). Column temperature was set to 45°C. Separation was achieved with a 110 min gradient (see Table 3.5) of solvent A (UPLC/MS grade water, Biosolve, 232141, with 0.1% TFA, ProteoChem, LC6203) and solvent B (UPLC/MS grade acetonitrile, Biosolve, 012041, with 0.1% TFA, ProteoChem, LC6203) at a flow rate of 350 nL/min. 7.9 µL (*in vitro* validation) or 7.3 µL (*in vivo* validation) of each

Time (min)	% A	% B
0	95	5
0.33	89	11
1	89	11
3.66	86	14
67.66	70	30
81	60	40
90	15	85
95	15	85
97	95	5
110	95	5

Table 3.5.: UPLC gradient used for discovery experiment. A step-wise increased, linear gradient of solvent B (acetonitrile, 0.1% TFA) in solvent A (water, 0.1% TFA) was applied for peptide elution.

sample were injected by partial loop injection and loaded on the column within 40 min (5% solvent B). Samples were separated into 1200 fractions by spotting on a Opti-TOF™ LC/MALDI target plate (AB SCIEX, 1018497) using a SunCollect MALDI Spotter (SunChrom) with 10 min delay time, 4 s per spot and 2 µL/min matrix flow rate. Eluted fractions were thereby robotically mixed with α -cyano-4-hydroxycinnamic acid (CHCA) matrix (3 mg/mL, in 80% acetonitrile, 0.1% TFA in water, ProteoChem, p9100) combined with four internal standard peptides (see Table 3.6). Standard peptide mixture was prepared out of 0.333 nmol/µL stock solutions of the single peptides in 50% acetonitrile, 0.1% TFA in water, aliquoted and dried in a vacuum concentrator to be mixed with 2 mL CHCA matrix directly before use. MALDI target plates were stored dry at -20°C for subsequent MALDI-MS analysis.

Peptide Sequence	[M+H] ⁺	c (fmol/μL)	fmol/spot
EEQPSTPAPKVEQQEEILC	2155.0231	300	40
CLEHMYHDLGLVRDF	1846.8732	150	20
TGVFDEAIRTGVF	1411.7221	75	10
TVFDEAIR	951.0688	37.5	5

Table 3.6.: Internal standard peptides. Standard peptides were spiked into CHCA sample matrix in the given concentration. Per spot, 133 nL were mixed with eluted peptide fractions (23 nL) before being deposited on a MALDI target plate.

For *in vitro* validation, peptides derived from a total protein amount of 500 μg were subjected per LC/MS run. Kidney samples were derived from 750 μg, liver samples from 1.25 mg total protein amount per LC/MS run.

3.7.1.2. MS Acquisition (MALDI-TOF System)

Mass spectrometric analysis was performed on a MALDI TOF/TOF™ 5800 system (AB SCIEX) using the TOF/TOF™ Series Explorer™ Software (AB SCIEX, version V4.1.0) with oracle database schema version 4.0.0, data version 4.0.5.

MS1 spectra acquisition was performed with *MS Reflector Positive* as operating mode within the MS acquisition method *MS_reflector* (or *MS_reflector_1000Shots*). Mass range was set to 750-4000 Da (focus mass 2000 Da). 250 shots per subspectrum were acquired, 8 (4) subspectra were accumulated (total: 2000 (1000) shots per spectrum). During acquisition, the sample plate was moved with continuous stage motion with a stage velocity of 1000 μm/s. The digitizer used 0.5 ns bin size with a vertical scale of 0.5 (V full scale), a vertical offset of -0.5 (% full scale) and an input bandwidth of 1000 MHz. Laser pulse rate was set to 400 Hz. Detector voltage multiplier was always adjusted following general maintenance of the instrument. The Y2 and X2 deflector voltages were manually adjusted before sample measurement using the MS processing method *MS_Reflector_Internal_new* and five internal standard peptides as reference masses, that were deposited manually on calibration spots (see Table 3.7). The reference peptides were prepared by mixing one Peptide Calibration Standard Set vial (ProteoChem, s6104) with one ACTH 18-39 vial in UPLC-grade water to a stock solution with a final concentration 200 μg/mL per peptide. The stock solution was stored and twenty-fold diluted with CHCA matrix solution before use. For calibration, 0.5 m/z were taken as mass tolerance, maximal outlier error was 10 ppm. The laser intensity was adjusted on a sample spot using the MS processing method *MS_Reflector_3IntPeptides* and the internal standard peptides as reference masses (see Table 3.6). For sample measurements, the MS processing method *MS_Reflector_3IntPeptides* defined the following peak detection characteristics: Minimal S/N was set to 5, the local noise window width to 250 m/z and the minimal peak width at the FWHM to 1 bin. Raw spectrum filtering contained peak smoothing with FFT and Poisson Denoise.

Protein name	Peptide Sequence	[M+H] ⁺	c (pmol/μL)
Leu-Enkephalin	YGGFL	556.277	18
Gonadoliberin	pEHWSYGLRPG	1182.581	8.4
Angiotensin I	DRVYIHPFHL	1296.685	7.7
Neurotensin	pELYENKPRRPYIL	1690.928	5.9
ACTH 18-39	RPVKVYPNGAEDESAEAFPLEF	2465.199	4.0

Table 3.7.: Standard peptides for calibration. Standard peptides were spiked into the CHCA matrix in the given concentration. 0.8 μL per spot were manually deposited on the eight calibration spots on the MALDI target plate.

Monoisotopic peaks were flagged using H adduct and the generic formula C₆H₅NO. Cluster area S/N optimization was performed with a S/N threshold of 20. For internal calibration, the four internal peptides spiked into the sample matrix were used as reference masses (see Table 3.6). Only monoisotopic peaks were used, 0.3 m/z were taken as mass tolerance, maximal outlier error was 10 ppm. After sample plate alignment, MS analysis was submitted as batch acquisition with plate model and default calibration.

MS/MS spectra acquisition was performed with *MS-MS 1KV Positive* as operating mode within the MS/MS acquisition method *MSMS_DynamicExit*. Collision induced dissociation and automatic acquisition control were activated. Metastable suppressor was selected, 250 shots per sub-spectrum were acquired and a maximum of 12 subspectra were accumulated (total: 3000 shots per spectrum) or stopped when the final spectrum reached the desired quality (high). During acquisition, the sample plate was moved with continuous stage motion with a stage velocity of 1200 μm/s. The digitizer used 1.0 ns bin size with a vertical scale of 0.5 (V full scale), a vertical offset of 0.1 (% full scale) and an input bandwidth of 200 MHz. Laser pulse rate was set to 1000 Hz. Detector voltage multiplier was always adjusted following general maintenance of the instrument. Y2 and X2 deflector voltages, TIS offset and the laser intensity were manually adjusted before sample measurement using the MS/MS acquisition method *MSMS_Tune* (total: 2000 shots per spectrum) and the MS/MS processing method *MSMS_Positive_Internal_Cal_2465* using the 1326.6800 m/z fragment of the ACTH peptide as reference mass, that was deposited manually on calibration spots (see Table 3.7). 0.4 m/z were taken as mass tolerance, maximal outlier error was 0.05 m/z. For sample measurements, the processing method *MSMS_Default* defined the following peak detection characteristics: Minimal S/N was set to 15, the local noise window width to 250 m/z and the minimal peak width at the FWHM to 1.5 bins. Raw spectrum filtering contained Savitsky-Golay smoothing of peaks with 5 points across the peak and polynomial order 4. Monoisotopic peaks were flagged using H adduct and the generic formula C₆H₅NO. Cluster area S/N optimization was performed with a S/N threshold of 15. The interpretation method for sample runs triggering the acquisition and processing method was *LC_Precursor_Selection_Dynamic_Exit* (or *LC_Precursor_Selection_Dyna-*

mic_Exit_StrongestFirst, used for tissue samples) with a mass range of 750-4000 Da and a minimum S/N filter of 50. A maximum of 35 precursors per fraction was selected (maximum of 40 000 per LC run) and, based on MS1 intensity, weakest (or strongest) precursor was acquired first. After sample plate alignment, MS/MS analysis was submitted as batch acquisition.

3.7.2. SRM Analysis

3.7.2.1. Peptide Selection

Based on the SRM Atlas data (www.srmatlas.org, March 2015) for human and mouse, up to five proteotypic peptides per target protein were selected for selected reaction monitoring (SRM). Mouse-specificity and proteotypicity of the selected candidate peptides as well as the existence of a tryptic cleavage site before the target sequence was tested using PepSir (see chapter 3.8.3) and the NCBI Protein Blast[®] (<http://blast.ncbi.nlm.nih.gov/Blast.cgi>) with peptide sequence information from UniProt database (www.uniprot.org, March 2015). Peptides were restricted to a mass range of 600-2000 Da, methionine and cysteine containing peptides were excluded if possible. To determine the levels of the endogenous target peptides in kidney and liver mouse samples, heavy peptide standards corresponding to their natural counterparts (light) were synthesized with heavy isotopic lysine ($^{13}\text{C}_6^{15}\text{N}_2$) or arginine ($^{13}\text{C}_6^{15}\text{N}_4$) at the C-terminus (Intavis, Heidelberg, Germany), pooled and spiked into the digested and desalted samples to a final concentration of 0.5 pmol/ μL .

3.7.2.2. UPLC-Based Separation

Samples subjected to SRM analysis were separated by a nanoACQUITY UPLC system (Waters) equipped with a nanocapillary, reversed-phase M-Class Peptide CSH C18 Column, 130 Å pore size, 1.7 μm particle size, 300 μm x 150 mm (Waters, 186007563). Column temperature was set to 55°C. Separation was achieved with a 120 min gradient (see Table 3.8) of solvent A (UPLC/MS grade water, Biosolve, 232141, with 0.1% FA, 0.01% TFA, ProteoChem, LC6202) and solvent B (UPLC/MS grade acetonitrile, Biosolve, 012041, with 0.1% FA, 0.01% TFA, ProteoChem, LC6202) at a flow rate of 6 $\mu\text{L}/\text{min}$. 10 μL of each sample were injected by full-loop injection (overfill factor 1.3, kidney samples or 1.1, liver samples) and loaded on the column within 4 min (3% solvent B).

For *in vivo* kidney validation, peptides derived from 1.5 mg total protein amount were subjected per SRM experiment and peptides from 1.4 mg total protein amount for the liver samples. For the full proteome kidney and liver samples, peptides derived from 40 μg precipitate were analysed per run.

3.7.2.3. SRM Acquisition

SRM analysis was performed on a QTRAP[®] 6500 system (AB SCIEX) operated with Analyst[®] software (version 1.6.2 build 8489). For peptide identity confirma-

Time (min)	% A	% B
0	97	3
1	96	4
110	70	30
111	15	85
114	15	85
115	97	3
119	97	3

Table 3.8.: UPLC gradient used for SRM experiment. A step-wise increased, linear gradient of solvent B (acetonitrile, 0.1% FA, 0.01% TFA) in solvent A (water, 0.1% FA, 0.01% TFA) was applied for peptide elution.

tion and SRM method optimization/validation, IDA runs were performed on the heavy peptide pool. Using the positive mode, *Enhanced MS* spectra were acquired with a scan speed of 10 000 Da/s in a mass range of 400-2000 Da, followed by *Enhanced Resolution MS* spectra acquisition of the three most intense peaks in the individual spectra with a scan speed of 50 Da/s. Selection criteria included only ions in a mass range of 450-2000 Da with charge state +2 or +3. Former target ions were excluded for 1 min after 15 occurrences. Rolling collision energy was used, mass tolerance was set to 250 mDa, isotopes within a 4 Da window were excluded. Enhanced resolution scans were used to confirm charge states by the isotope pattern. MS/MS spectra were acquired in the trap mode (*Enhanced Product Ion*) with dynamic fill time, a mass range of 100-2000 Da and a scan speed of 10 000 Da/s. To confirm peptide identity, .wiff files were searched in the ProteinPilot™ software (see chapter 3.8.2) against a database containing the .fasta file information for the target proteins spiked into an *E.coli* background (4303 entries, downloaded 2013/08/01 + 39 entries, downloaded 2015/04/02). The best three transitions for each peptide were selected using the Skyline software (version 2.6.0.6851) based on maximum signal intensities. Final transition lists were created by selecting two peptides per protein based on signal intensities, peptide length, cysteine content and retention time. Retention times were adapted by a scheduled SRM measurement of the heavy peptides in a real sample background.

For the final SRM quantification experiment, two reproducibly detectable peptides per protein with at least 2 charges were targeted with three SRM transition signals per heavy or light peptide (see Suppl. Tables B.1, B.2). This resulted in a total of 384 (252) transitions for 64 (42) peptides deriving from 32 (21) proteins for the kidney (liver) dataset acquired. Scheduled SRM analyses were performed with Q1 operated in unit resolution, Q3 in low resolution, a target scan time of 2 s, a median (minimal) dwell time of 106 ms (34 ms) and retention time windows of ± 2 min around the specific elution time within the kidney dataset and a median (minimal) dwell time of 162 ms (43 ms) and retention time windows of ± 2.25 min around the specific elution time within the liver dataset.

3.8. Analysis of Mass Spectrometric Data

3.8.1. Data Extraction and Generation of 2D Peptide Maps

1200 MS spectra (.t2d files) per sample run were extracted using a macro running in the Data Explorer® Software (Applied Biosystems/MDS Analytical Technologies, Version 4.10 build 124), resulting in the creation of a Peak and a Spectra folder. For visualisation of the UPLC based separation and the MS measurements, 2D-Peptide maps were generated from the spectra files using 2D Peptide Map (version 0.1.5.0) created by Dr. Tim Fugmann in Microsoft Visual Basics. Settings used included 1411 m/z as internal standard peptide peak (see Table 3.6), normalization to a standard peak height of 1000 or 500 and peak height capping at 100, 500 or 1000, depending on spectral intensity.

3.8.2. Peptide and Protein Identification

Spectra from MS/MS analysis were processed with the ProteinPilot™ Software (AB SCIEX, version 4.5, revision 1656). Searches for peptide and protein identification were performed using the Paragon™ algorithm (AB SCIEX, version 4.5.0.0, 1654). The MALDI TOF/TOF™ 5800 system was selected as instrument, iodoacetamide as cysteine modification and trypsin as digestion enzyme. ID focus included biological modifications and amino acid substitutions. Search effort was set to Thorough ID, a protein confidence score of 1.3 (95%) was used as threshold and a false discovery rate (FDR) was calculated for every search. Depending on sample type, spectra were searched against a human (70101 entries, downloaded 2012/07/30) or mouse database (51514 entries, downloaded 2012/08/08) downloaded as .fasta files from www.uniprot.org. Exported peptide summaries were used for further analysis.

3.8.3. Determination of Peptide Proteotypicity

Peptide Summaries from the ProteinPilot™ search were dragged and dropped into PepSir (version 1.6.1), a software tool for identification of proteotypic peptides with respect to a given database, created by Dr. Alexander Kerner.^[282] Depending on sample type, files were searched against a human (68485 entries, downloaded 2015/05/11) or mouse database (44628 entries, downloaded 2015/01/26) without isoforms downloaded as .fasta files from www.uniprot.org. Protein/Gene Name was used as identifier of proteotypicity if a peptide sequence was found more than once. Filtering was set to Filter by Confidence and ignore below 95%. Peptides classified as not proteotypic (*degenerated*) or *not found* were automatically removed from the resulting peptide summary. The summaries of proteotypic peptides were either analysed with respect to the number of identified proteotypic peptides/proteins in different samples or used as part of the input file to MS_QBAT for propagation of annotations (see chapter 3.8.5) or relative quantification (see chapter 3.8.4).

3.8.4. Relative Quantification Using MS_QBAT

The peptide summary of proteotypic peptides (see chapter 3.8.3) was included in the corresponding peak file folder of the extracted MS1 data (see chapter 3.8.1) and dragged and dropped into MS_QBAT software (version 2.2.1) for MS1 based, label-free quantification created by Dr. Alexander Kerner.^[282] To compensate for run-to-run variability between different samples, a fraction-wise intensity normalization against the four internal standard peptide peaks (see Table 3.6) identified with a tolerance range of ± 0.5 Da was performed as a first step. Secondly, the normalized MS1 peaks were annotated based on m/z and retention time using the information from the peptide summaries of proteotypic peptides (MS/MS data). Hereby, a fraction delta of ± 1 and a m/z delta of 50 ppm was used; fraction length was set to 4 s, column identifiers for m/z, sequence and protein ID were given in accordance to the peptide summary headers. Next, individual features were extracted. Feature identification was based on (i) m/z tightening with a mass shift of 50 ppm threshold, (ii) fraction tightening with a gap size of maximal 5 fractions (6 for *in vivo* validation of liver samples) and (iii) annotation tightening to ensure that there is only one peptide identification per feature. To obtain a more stable quantification by reduction of noise, the total feature number per sample was drastically reduced with a complexity reduction step: Applied S/N filter as well as feature length filter were automatically optimized in a range of 0-400 in intervals of 10 and 0-4 in intervals of 1, respectively, so that at least 99.5% of the identifications were kept.

To map features between different sample runs, feature alignments of pairs of two samples (cross-alignments) were performed with a maximal fraction delta of ± 20 and a maximal mass delta of ± 10 ppm during mapping. Alignments with a difference of more than ± 4 to the alignment path, that maps the retention times from sample A to retention times from sample B, were orphanized. In the next step, the alignments of feature pairs were used to propagate the annotation information between two samples, if a feature was not annotated in one sample, but in the other. Hereby, annotations were only propagated between pairs of samples treated with the same biotinylation reagent.

An alignment by a genetic algorithm (GEAL alignment) was used to find the best parameters for the alignment of single samples or super-samples (GEAL alignment of several samples combined in one file) against each other, which is the basis for relative quantification. 200 iterations were performed per GEAL alignment with the values given for the cross-alignment as starting values. Alignment settings included a fraction range of 10-200 (interval 10), a m/z range of 10-200 ppm (interval 10) and an alignment path range of 2-60 (interval 2). If the annotation information of two aligned features did not match, the alignment was not considered for quantification. Annotations only present in one sample were automatically transferred to the other sample as described in the annotation propagation step. In presence vs. absence situations, MS_QBAT replaced the non-existing value with an introduced background intensity value to enable relative quantification. If more than two samples were aligned to each other, MS_QBAT automatically calculated a sample similarity score

before starting the GEAL alignment with the samples of highest similarity.

GEAL alignments were used to quantify proteins between pairs of technical or biological replicates as well as between pairs of super-samples biotinylated with different reagents. Protein quantification was based on the ratio between the summed peptide intensity values belonging to this protein. P-values for the single proteins were calculated with an unpaired t-test based on peptide intensity values. Quantification tables containing information about protein ID and intensity values per sample, intensity ratio, p-values and the number of peptides used for the quantification were exported using the raw intensity values.

For quantification against another sample type (biotinylated with a different reagent), technical or biological replicates were combined to super-samples by GEAL alignment. If technical and biological replicates were present within the same dataset, the alignment was performed in a two-step procedure: Super-samples of technical replicates were aligned to a combined super-sample of all biological replicates. For the *in vitro* validation experiment (see chapter 3.6.1), technical triplicates per biotinylation reagent were combined to super-samples. For *in vivo* validation (see chapter 3.6.2), two technical and four biological replicates per biotinylation reagent were combined to super-samples in the kidney dataset; and three biological replicates were combined in the liver dataset. All samples were quantified against the super-sample from biotinylation with the commercial reagent Sulfo-NHS-LC-biotin.

3.8.5. Propagation of Annotations Using MS_QBAT

To account for mass spectrometric performance differences during optimization of the sample preparation procedure, annotations were transferred between comparable samples biotinylated with the same reagent. The peptide summary of proteotypic peptides (see chapter 3.8.3) was included in the corresponding peak file folder of the extracted MS1 data (see chapter 3.8.1) and dragged and dropped into MS_QBAT software (version 2.2.1). The following steps have been performed as described in chapter 3.8.4: annotation, feature extraction, cross-alignment of all possible sample pairs and propagation of annotations. The resulting MS_QBAT files were saved and analysed with respect to the number of identified proteotypic peptides/proteins in different samples.

3.8.6. Analysis of Protein Localization

For analysis of protein localization, subcellular localization information was extracted from www.uniprot.org using the NextDataTool (versions 2.1.3, 2015/02/25 and 2.1.4, 2015/05/20) created by Dr. Alexander Kerner. Localization information was manually curated using the following categories: plasma membrane (pm), extracellular and secreted (ex), intermembrane or cytoplasmic side of plasma membrane (ipm), membrane associated (mem), intracellular (int), no localization information found (no).

3.8.7. Analysis of SRM Data

SRM data was processed using the Skyline software (MacCoss Laboratory, version 2.6.0.6851). For correct peak detection, the default peak boundary assignment based on Savitzky-Golay smoothing was manually reassigned if required. Peptides with unfavorable elution profile or interfering peaks in the light transitions were excluded from further data analysis. To exclude peptides exhibiting only background noise in the light transitions leading to instable background quantification, the peak areas of the three measured transitions per peptide were normalized to the summed peak area (100%) within the Skyline software. Only peptides exhibiting a stable distribution of the transition ratios in the light transitions among all samples but the negative controls were included in further data analysis.

Information including background-reduced peak area of heavy and light peptides were exported for further analysis. For each peptide, peak areas of corresponding transitions were summed. The ratio between the background reduced peak area of the light transition and the background reduced peak area of the heavy transition was calculated to correct for ionization or spray differences between runs.

To evaluate, if the signals of the light transitions are within the background and therefore not suitable for a proper quantification, the average background-reduced peak area across all biotinylated surface enriched samples or across all full proteome samples was compared to the background-reduced peak area of the light peak background signal of heavy peptide pool in solvent. To obtain a reliable quantification, peptides with an average peak area less than four-fold up-regulated compared to the background signal were excluded from analysis. In the following, peak areas were normalised to the maximal peak area achieved among the enriched surface samples for this specific peptide. Mean and standard deviation of four (kidney dataset) or three (liver dataset) biological replicates were calculated per reagent group or full proteome sample.

3.8.8. Further Data Analysis Tools

Raw data of the mass spectrometric datasets were analysed with Microsoft Excel 2007. Statistical analyses and data visualisation was performed with GraphPad Prism (version 6.05).

Enrichment analysis of cellular component and molecular function was performed by gene ontology (GO) analysis using the WEB-based GEne SeT AnaLysis Toolkit (Webgestalt).^{[283],[284]} *Homo sapiens* or *mus musculus* was selected as organism, data were loaded with *uniprot_swissprot_accession* selected as ID type. GO analysis was performed with the organism's genome sets as reference for the enrichment analysis. Data were visualized as directed acyclic graph.

Cluster analysis was performed using the data visualisation platform GENE-E.^[285] *City-block distance* was used as distance metric, *average linkage* as linkage method. Color settings were customized using a *global* color scheme with 4-fold up-regulation depicted in red and 4-fold down-regulation in blue.

Protein abundance information was extracted from PaxDB, an absolute protein abundance database containing protein abundance data across various organisms and tissues.^{[286],[287]} The abundance information is based on the integration of several publicly available experimental datasets via a spectral counting pipeline.

The ProtParam tool was used to compute protein parameters such as protein or peptide mass and the amino acid composition.^[288]

Chemical structures were visualized with ChemBioDraw Ultra (version 14.0.0.117). 3D models were generated in Chem3D Pro (version 14.0.0.117).

4. Results

4.1. Biotinylation Reagents: Chemical Entities

4.1.1. Rationales and General Design of Biotinylation Reagents

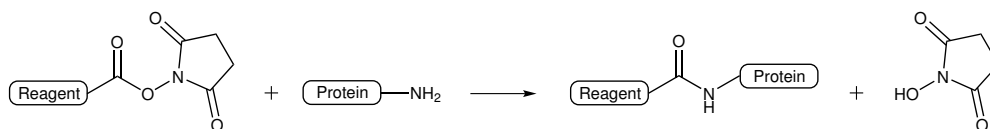
The identification of vascular accessible, disease-specific biomarkers is a prerequisite for the development of novel diagnostics or targeted therapies with fewer effects in healthy tissues. Hereby, the molecules of interest, targetable with specific monoclonal antibodies or antibody-drug conjugates, are proteins expressed on the cell surface of diseased cells, on endothelial cells of newly formed blood vessels in the diseased tissue or in the perivascular extracellular matrix.^{[160],[289]} Global proteomic approaches for the identification of novel biomarkers have to deal with low expression levels of the highly heterogeneous surface proteome fraction compared to the intracellular fraction as well as with their hydrophobic character.^{[146],[120]} The nature of a membrane protein is determined by hydrophobic membrane interacting domains on the one hand and by hydrophilic protein regions as well as polar posttranslational modifications like glycosylation on the cell surface on the other hand.^{[146],[161]} Peripheral membrane proteins can be associated with the cell membrane by interaction with integral proteins or the lipid bilayer itself via hydrophobic regions, whereas integral membrane proteins contain 1-20 transmembrane domains consisting of 15-25 hydrophobic amino acids each.^{[133],[120]} Therefore, besides the low abundance, the main challenges in membrane proteomics are the prevention of protein aggregation and efficient solubilisation with detergents compatible with LC/MS analysis, as well as the lack of polar arginine or lysine residues for tryptic cleavage within former membrane-spanning domains complicating the generation of short peptide fragments for proper mass spectrometric analysis.^{[290],[119]}

Enrichment of the proteome fraction of interest prior to analysis reduces sample complexity and enables a stable identification and quantification of potential biomarkers. One avenue towards the enrichment of vascular accessible, surface biomarkers is their modification with membrane-impermeable biotin derivatives followed by protein extraction and streptavidin-based enrichment on a solid support before proteomic analysis. A biotinylation reagent for enrichment of vascular accessible proteins consists of three building blocks:

- (i) A biotin moiety used for streptavidin-based enrichment.
- (ii) A linker determining the properties of the biotinylation reagent.
- (iii) A reactive moiety for covalent labelling of target proteins.

The complex formed by streptavidin homotetramers and biotin exhibits a very strong interaction ($K_d \sim 10^{-15}$ M)^[291] and has therefore been widely exploited for affinity-based detection or capturing of biotinylated target molecules.^[292] Whereas the tetrahydroimidizalone ring is of greatest importance for the binding to the avidin site, the terminal carboxylic group at the valeric acid side chain can be used for coupling of the biotin residue to another molecule.^{[291],[191]}

Proteins are composed of a variety of functional groups exploitable for chemical derivatization: Carboxylic groups react with amide-bond forming reagents upon activation, primary amino groups can be derivatized with activated ester compounds, secondary amino groups with tetra- or pentafluorophenyl ester derivatives. Thiol groups can be addressed with maleimido, haloacetyl or thiol groups to form thioether or disulfide bonds. Aldehyde groups generated by periodate oxidation of N-terminal serine or attached carbohydrate residues can be targeted with hydrazide compounds. Phenolic side chains of tyrosine residues can be alkylated, acylated or can undergo electrophilic addition.^{[191],[293]} For successful labelling of the whole cell surface proteome fraction, the targeted functional group has to be expressed in high frequency and in an accessible form on the cell surface. Thiol groups in cysteine residues are rare and often occurring in oxidized form as disulfide bonds stabilizing the protein's tertiary structure.^{[293],[294]} Activation of the carboxylic acid side chains of glutamic or aspartic acid or the protein C-terminus would lead to a certain degree of protein polymerization on the cell surface as both carboxylic and amino groups are present. Therefore, the highly abundant, primary amino groups occurring as ϵ -amines in lysine side chains or as α -amine at the protein N-terminus are the coupling sites of choice for stable labelling with biotin ester derivatives in a one-step reaction (see Scheme 4.1), although some lysine residues are lost as tryptic cleavage sites within the protein.



Scheme 4.1.: Modification of protein amino groups (ϵ -amines in lysine side chains or N-termini) with NHS-ester derivatives.

As outlined above, the linker between the biotin moiety for affinity purification and the reactive group for protein labelling determines the properties of any biotinylation reagent. First of all, the linker serves as a spacer to avoid steric hindrance during labelling of the bulky protein with reagent molecules as well as during the capturing process on the streptavidin resin. In order to allow for the biotinylation of proteins on living cells, the reagent has to be soluble in aqueous buffers at reasonable concentrations. Furthermore, the reagent must not cross biological membranes due to the localization of potential vascular accessible biomarkers on the surface of cells. To maintain plasma membrane integrity, the cell viability should not be affected during the biotinylation procedure.

4.1.2. Commercial Biotinylation Reagents and Advancements

Sulfo-NHS-LC-biotin (see Figure 4.1a) is one of the most popular biotinylation reagents available on the market. Sulphation of the ester group supports the water solubility of the molecule, but the 8.7 Å LC (long chain) linker between the biotin moiety and the activated carboxylic group still determines the clear hydrophobic character of the spacer arm. In other commercial compounds the aliphatic chain is substituted with a more hydrophilic polyethylene glycol (PEG) linker available in different spacer lengths up to 45.9 Å (see Figure 4.1b). Within this thesis, all validation experiments of the novel biotinylation reagents are performed against the commercial reagents Sulfo-NHS-LC-biotin and NHS-PEG₁₂-biotin.

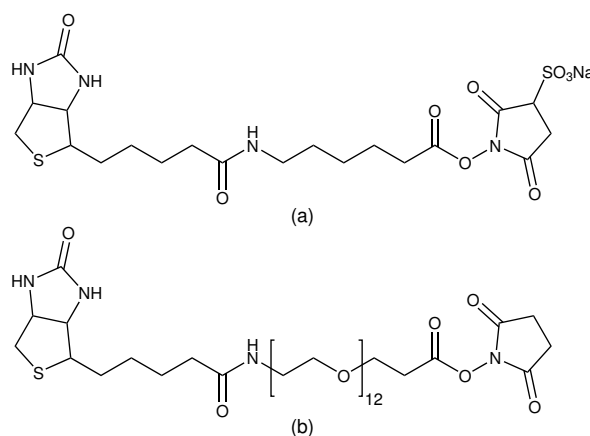


Figure 4.1.: Commercial biotinylation reagents used for various validation experiments. (a) Sulfo-NHS-LC-biotin, MW 556.58 g/mol; (b) NHS-PEG₁₂-biotin, MW 941.10 g/mol.

The analysis of whole cell lysate samples without prior enrichment yields approximately 10% of all identified proteins to be annotated to the plasma membrane.^{[130],[180]} By biotinylation reagent based enrichment, both the total number of identified surface proteins as well as the surface proteome fraction are increased to about 30%.^[180] The analysis of datasets from *in vivo* biomarker discovery studies performed with the commercial Sulfo-NHS-LC-biotin also confirms the identification of a considerable proportion of proteins not annotated to the cell surface or the extracellular matrix.^{[186],[295],[296]} Explanations for this phenomenon include (i) plasma membrane crossing of the reagents via diffusion or biotin transporters, (ii) co-purification of intracellular proteins attached to the biotinylated membrane proteins, (iii) wrong or incomplete GO annotation of the protein localization within the database as well as (iv) the presence of a fraction of apoptotic cells with increased membrane permeability.^[12]

In 2010, the novel biotinylation reagent NHS-β-Ala-(L-Asp)₃-biotin was published (see Figure 4.2).^[12] Compared to commercial Sulfo-NHS-LC-biotin, the reagent is highly water soluble due to three negatively charged aspartic acid side chains and exhibits a 36% increase in molecular weight. A comparative proteomic analysis of the top 50 regulated proteins of NHS-β-Ala-(L-Asp)₃-biotin or Sulfo-NHS-LC-biotin

perfused mouse kidney and liver tissue revealed a clear up-regulation of surface or extracellular matrix annotated proteins with the novel reagent.^[12] Nevertheless, only 364 proteins (kidney) and 392 proteins (liver) with a surface proteome fraction of 60% and 42%, respectively, were identified. The poor reactivity of the NHS- β -Ala-(L-Asp)₃-biotin is further elucidated in chapter 4.1.4.

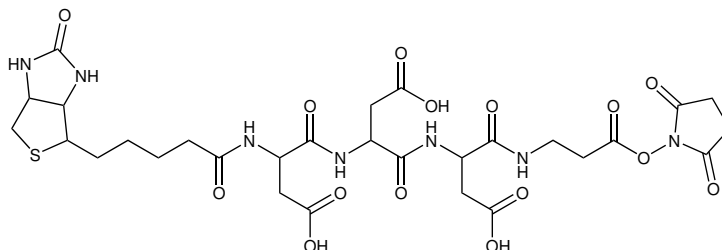


Figure 4.2.: NHS- β -Ala-(L-Asp)₃-biotin, MW 757.73 g/mol published by Strassberger *et al.* in 2010 with optimal NHS-activation.^[12] See Figure 4.5 for 3D model.

4.1.3. Analysis of the Biotinylation Procedure

The reactivity of different biotinylation reagents as well as particular labelling conditions were assessed with BSA as model protein. BSA contains 60 primary amino groups (59 lysine side chains and the N-terminus). Biotinylation reagents were dissolved or diluted in aqueous buffer directly before use and were incubated with 1 eq BSA in different ratios (1 eq - 100 eq) for 15 min at room temperature in PBS buffer (if not indicated differently). Mass spectrometric analysis of the desalted BSA samples was performed on a MALDI TOF/TOFTM 5800 system (AB SCIEX). The difference between the m/z at the peak maximum of the singly charged peak to the m/z of unmodified BSA was used to calculate the number of reagent tags added to one BSA molecule.

The labelling reaction occurs very fast: A prolongation of the biotinylation reagent's incubation time with the target protein from 30 s to 1 h did not change the degree of modification (see Figure 4.3a). Hydrolysis of the NHS ester always competes with the favored reaction of the primary amino group, because the labelling reaction has to be performed in aqueous buffer when working in an *in vitro* or *in vivo* system. Therefore, biotinylation reagents have to be stored in dried form or in a water-free solvent and have to be dissolved or diluted directly before use. Decreasing the temperature of the biotinylation solution also decreases the hydrolysis rate (see Figure 4.3b). The coupling reaction is very efficient at physiological to slightly acidic pH, whereas the hydrolysis rate is heavily increased at basic pH (see Figure 4.3c). Batch-to-batch variations occur not only with freshly synthesized, but also with commercial reagents (see Figure 4.3d), most likely due to slightly different storage conditions leading to varying water content. Due to the fast reaction with primary amino groups, a non-reacted reagent excess can be efficiently quenched with an amine containing solution, such as Tris buffer.

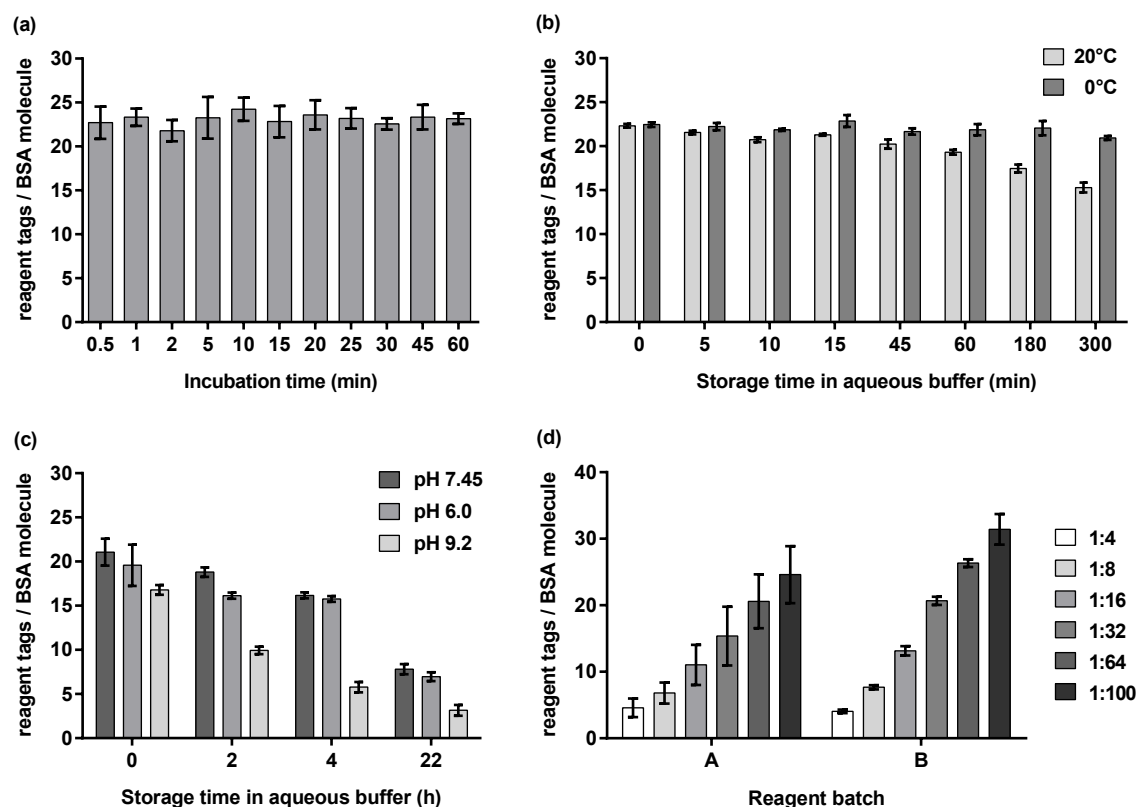


Figure 4.3.: Biotinylation of model protein BSA with Sulfo-NHS-LC-biotin under different conditions, analysis with linear MALDI-MS (mean and standard deviation, six to eight data points per condition). (a) Variation of incubation time with BSA, biotinylation ratio 1:64; (b) storage in PBS at 0°C or 20°C before incubation with BSA, biotinylation ratio 1:64; (c) storage in PBS (pH 7.45), MES (pH 6.0) or carbonate buffer (pH 9.2) at 20°C before incubation with BSA, biotinylation ratio 1:100; (d) reactivity comparison of different reagent batches.

4.1.4. Reactivity of NHS- and sNHS- β -Ala-(L-Asp)₃-biotin

NHS- and sNHS- β -Ala-(L-Asp)₃-biotin (see Figure 4.2) were synthesized to further assess the poor reactivity during the labelling procedure. Carboxylic group activation was performed with an excess of 1.3 eq of NHS or with the hydrophilic sNHS in the presence of carbodiimide according to the authors' instructions.^[12] As the activated reagent is applied on protein samples without further purification, a large carbodiimide excess during activation has to be avoided to prevent crosslinking of the target proteins, a possible side reaction. Therefore, activation was either performed with 1.1 eq 1-Ethyl-3-(3-dimethylaminopropyl)carbodiimide hydrochloride (EDC-HCl) or with a 3- to 10-fold excess of N,N'-dicyclohexylcarbodiimide (DCC) resin, which was subsequently separated from the reaction mixture. Both activation strategies proved to be comparable as examined in reactivity tests on BSA.

The reactivity assessment of NHS- β -Ala-(L-Asp)₃-biotin and Sulfo-NHS- β -Ala-

(L-Asp)₃-biotin on BSA clearly indicates the poor reactivity of the peptide reagent compared to commercial Sulfo-NHS-LC-biotin (see Figure 4.4). The significantly decreased biotinylation rate originates from the structural difference of the linkers between the biotin moiety and the activated carboxylic group: The aspartic acid side chains in the peptide reagent need more space than the simple aliphatic linker of Sulfo-NHS-LC-biotin. This is why the reaction with protein amino groups in bulky protein regions is sterically hindered. Furthermore, the NHS- or sNHS-activation of the peptide is not possible in a site-specific manner: Biotin-(L-Asp)₃-β-Ala not only contains the C-terminal carboxylic group, but also three aspartic acid side chains. Activation with the slight excess of 1.3 eq NHS or sNHS leads to an undefined product mixture with more or less sterically hindered structures around the ester group, depending on the modified carboxylic group (see Figure 4.5). A decreased reactivity on target proteins is the consequence.

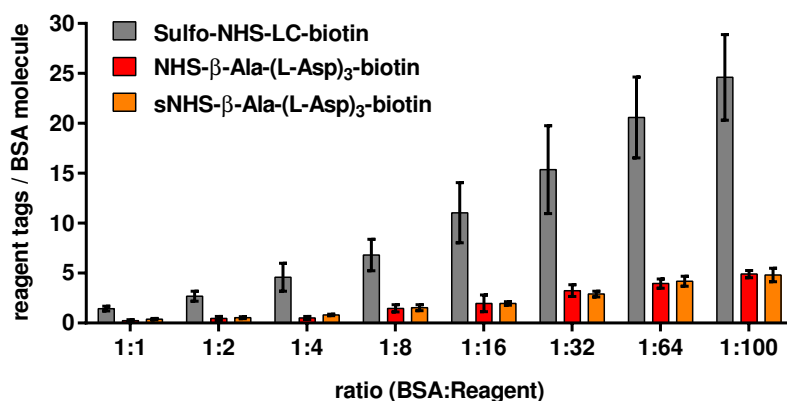


Figure 4.4.: Biotinylation of model protein BSA in different ratios with commercial Sulfo-NHS-LC-biotin and the peptide reagent NHS-β-Ala-(L-Asp)₃-biotin (Sulfo-NHS-β-Ala-(L-Asp)₃-biotin) published by Strassberger *et al.*^[12] Analysis with linear MALDI-MS (mean and standard deviation, seven to eight data points per condition).

4.2. Synthesis of Novel Biotinylation Reagents

4.2.1. Peptide-Based Biotinylation Reagents

4.2.1.1. Design and Synthesis of Peptide-Based Biotinylation Reagents

Two novel biotinylation reagents SMCC-Cys-(L-Asp)₃-biotin and SM(PEG)₆-Cys-(L-Asp)₃-biotin have been designed to improve the low reactivity of NHS-β-Ala-(L-Asp)₃-biotin, while keeping desired features such as the increase in size and polarity to decrease the membrane crossing potential compared to commercial reagents. A C-terminal thiol residue is introduced to the peptide linker and exploited for site-specific coupling to a bifunctional N-hydroxysuccinimide/maleimide crosslinker for

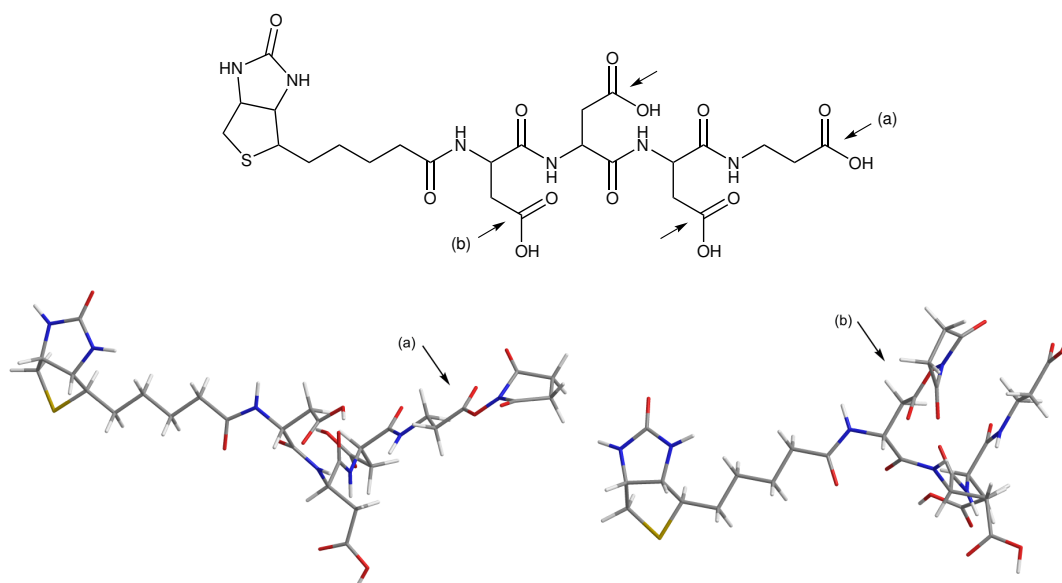
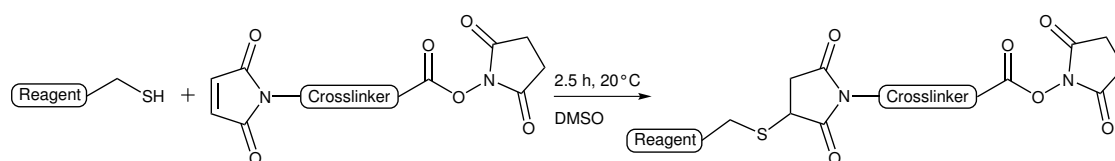


Figure 4.5.: Site-unspecific activation of Biotin-(L-Asp)₃-β-Ala. Arrows indicate possible positions for activation with NHS or sNHS. Activation at position (a) results in the least sterically hindered reaction product. 3D models of two exemplary product structures (activation at positions (a), (b)) demonstrate differential steric hindrance of the activated carboxylic group; atoms depicted in grey (C), white (H), red (O), blue (N), yellow (S).

activation of the reagent (see Scheme 4.2). The maleimide group efficiently undergoes alkylation with the peptide thiol group within 2.5 h at room temperature yielding a thioether linkage. The reaction efficacy could be improved by pre-treating the peptide with 0.5 eq TCEP for the reduction of disulfide bonds (to avoid the formation of peptide dimers and to generate free thiol groups). Subsequently, a slight excess of 1.3 eq succinimidyl-4-(N-maleimidomethyl)cyclohexane-1-carboxylate (SMCC) or succinimidyl-([N-maleimidopropionamido]hexaethyleneglycol)ester (SM(PEG)₆) crosslinker was added to the previously reduced peptide. As the bi-



Scheme 4.2.: Activation of novel peptide-based reagents by a bifunctional N-hydroxy-succinimide/maleimide cross linker.

functional crosslinkers already contain the activated NHS-ester group for protein amino group coupling, the reagent synthesis was performed in a water-free solvent to prevent hydrolysis. The structures of the novel peptide-based biotinylation reagents SMCC-Cys-(L-Asp)₃-biotin and SM(PEG)₆-Cys-(L-Asp)₃-biotin are depicted in Figure 4.6.

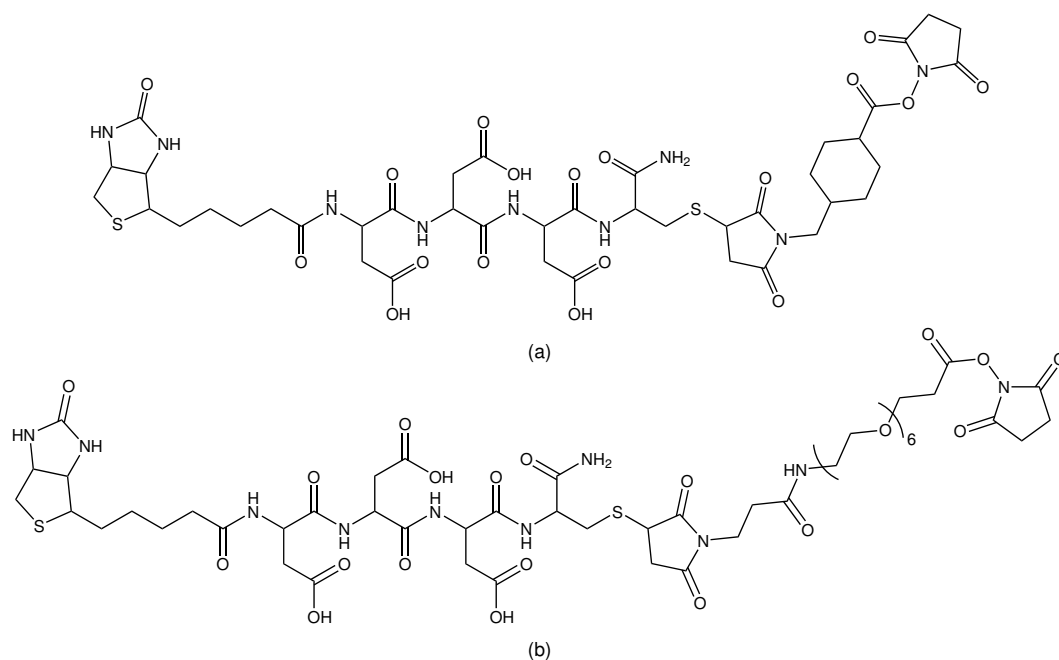


Figure 4.6.: Novel peptide based biotinylation reagents. (a) SMCC-Cys-(L-Asp)₃-biotin, MW 1026.06 g/mol; (b) SM(PEG)₆-Cys-(L-Asp)₃-biotin, MW 1293.33 g/mol.

4.2.1.2. Reactivity of Peptide-Based Biotinylation Reagents

For reactivity assessment on BSA as model protein, SMCC-Cys-(L-Asp)₃-biotin and SM(PEG)₆-Cys-(L-Asp)₃-biotin were compared with the commercial reagents Sulfo-NHS-LC-biotin and NHS-PEG₁₂-biotin as well as with the non-specifically activated reagent NHS-β-Ala-(L-Asp)₃-biotin (see Figure 4.7). The C-terminal, site-specific activation of the novel reagents significantly improves the reactivity compared to the non-specifically activated NHS-β-Ala-(L-Asp)₃-biotin, while the basic structure of the peptide linker is kept. Steric hindrance is drastically reduced, but cannot be fully avoided due to the more bulky peptide linker compared to the aliphatic (Sulfo-NHS-LC-biotin) or pegylated (NHS-PEG₁₂-biotin) linkers of the commercial reagents. The novel reagents are triply negatively charged at physiological pH and exhibit an 84-133% increase in size compared to Sulfo-NHS-LC-biotin. Due to the significant tag size, the resulting peaks in mass spectra of BSA biotinylated in high ratios are broadened, preventing a proper analysis of ratios above 1:64.

Different peptide synthesis batches have been used for the experiments performed within this thesis. Batch-to-batch variations (see Figure 4.8a) most likely arise from varying water content and pH in the lyophilized powder leading to different hydrolysis rates. The activated reagent undergoes a slight loss of activity upon storage in dry DMSO under inert gas (see Figure 4.8b) as hydrolysis cannot be fully prevented. All experiments within this thesis are performed with reagents that were directly activated before use.

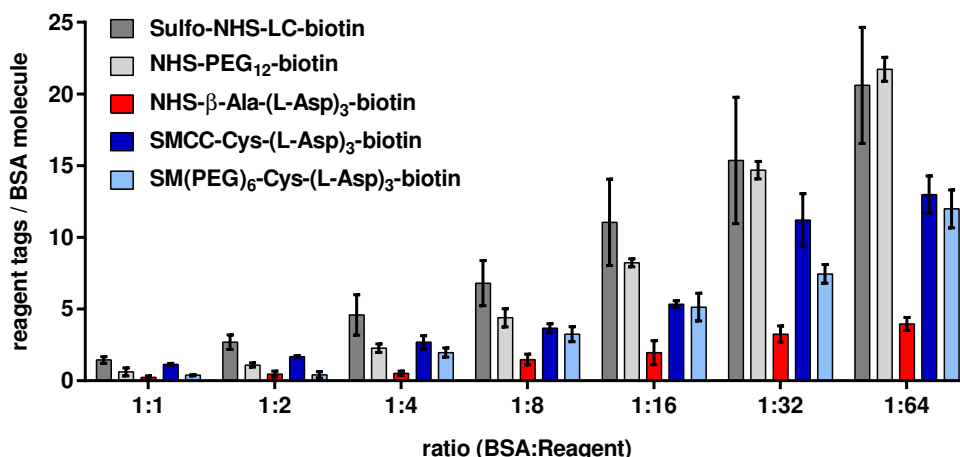


Figure 4.7.: Biotinylation of model protein BSA in different ratios, analysis with linear MALDI-MS (mean and standard deviation, five to eight data points per condition). The commercial reagents Sulfo-NHS-LC-biotin and NHS-PEG₁₂-biotin are depicted in grey, NHS-β-Ala-(L-Asp)₃-biotin published by Strassberger *et al.*^[12] in red, the novel reagents SMCC-Cys-(L-Asp)₃-biotin and SM(PEG)₆-Cys-(L-Asp)₃-biotin in blue.

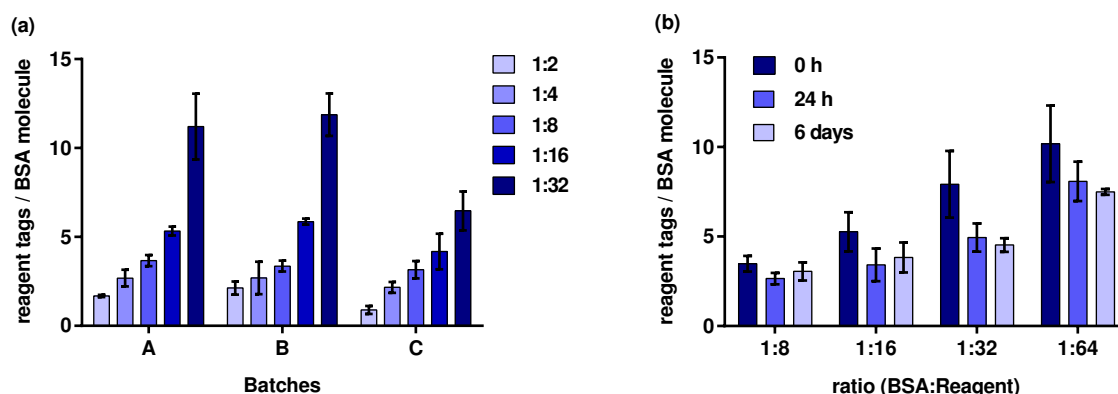


Figure 4.8.: Biotinylation of model protein BSA with SMCC-Cys-(L-Asp)₃-biotin, analysis with linear MALDI-MS (mean and standard deviation, up to eight data points per condition). (a) Reactivity of different peptide educt batches used for the validation experiments within this thesis; (b) reactivity loss upon storage in dry DMSO under inert gas at -20°C for 24 h or 6 days.

4.2.2. Heparin-Based Biotinylation Reagents

4.2.2.1. Design and Synthesis of Heparin-Based Biotinylation Reagents

Heparin is a mixture of linear glycosaminoglycans of different chain lengths composed of variable modified L-iduronic acid and D-glucosamine saccharide subunits joined through 1→4 glycosidic linkages. The exact composition is depending of the tissue type used for extraction of the heparin.^[297] The heparin sodium salt from porcine intestinal mucosa used within this thesis mainly consists of disaccharide

units composed of 2-O-sulfated iduronic acid and 6-O- and N-sulfated glucosamine (see Figure 4.9) and has a mass range of 13 000-17 000 g/mol with an average mass of 15 000 g/mol (25 disaccharide units). So-called low molecular weight (LMW) heparin can be obtained via alkaline depolymerization of heparin. The LMW heparin enoxaparin also used within this thesis originates from heparin of porcine intestinal mucosa and has an average mass of 4421 g/mol (8 disaccharide units). As a result of the high sulphation degree and the free carboxylic groups, the polymer is highly negatively charged in a wide pH range. Heparins are reported to still contain 2-10% water after extensive drying due to their extreme hydrophilicity and are practically insoluble in organic solvents.^[297]

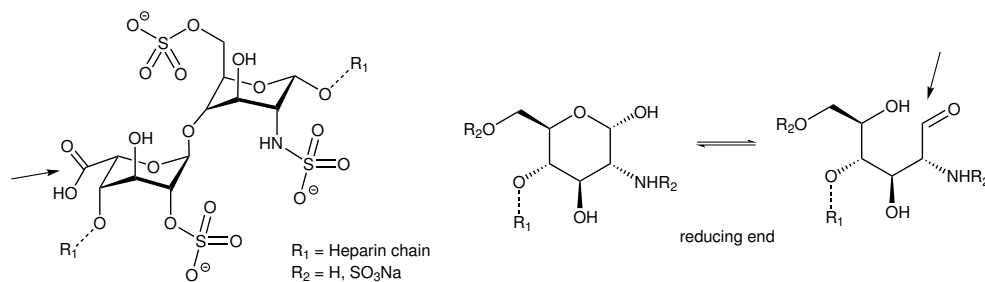
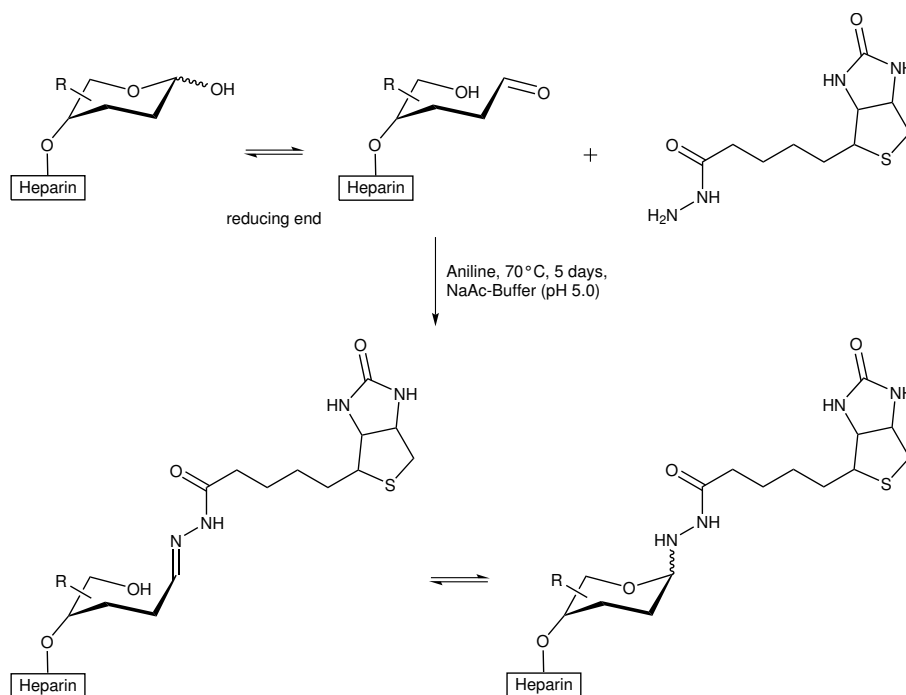


Figure 4.9.: Structure of heparin: the most common IdoA(2S)-GlcNS(6S) disaccharide subunit and the reducing end of the heparin chain. Arrows label the reactive groups used for synthesis of novel heparin-based biotinylation reagents.

The heparin chain incorporates a variety of functional groups exploitable for chemical derivatization. Every single heparin chain contains an intramolecular hemiacetal at the reducing end (see Figure 4.9), that is in equilibrium with its open-chain aldehyde form. Nucleophilic compounds such as hydrazides can target the anomeric centre.^{[298],[299],[300]} Moreover, each disaccharide subunit contains a free carboxylic group, several sulfo and hydroxyl groups, in some disaccharide units also as vicinal diols, as well as an average of 0.3 unmodified amino groups per heparin chain.^[297]

The large size, multiple charges, hydrophilicity and the various functional groups make the heparin structure to an ideal candidate for the synthesis of novel, membrane-impermeable biotinylation reagents.

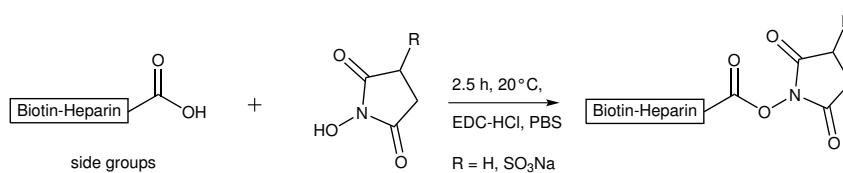
The reducing chain ends were exploited to introduce one biotin residue per heparin chain. The reaction rate of this process is low, as the equilibrium is in favor of the ring-closed form.^[301] Elevated temperature and an excess of nucleophile are required to drive the coupling reaction. Moreover, nitrogen nucleophiles more reactive than simple amines. Hydrazines, hydrazides or hydroxylamines are commonly used for modification.^{[293],[302],[303]} A pH range of 3.0 - 5.5 of the reaction buffer is reported to support favorable reaction kinetics.^{[302],[300],[304]} Here, a biotin hydrazide derivative was chosen for coupling in sodium acetate buffer at pH 5.0 (see Scheme 4.3). The reaction can be performed under reducing conditions, e.g. with sodium cyanoborohydride, to form a stable, reduced hydrazone linkage or under non-reducing conditions yielding hydrazone/glycosylhydrazone formation.^{[298],[299]} The latter is reported



Scheme 4.3.: Biotinylation of heparin at the reducing end.

to be stable against hydrolysis in a range from slightly acidic to basic pH.^{[305],[306]} The reaction rate can be further increased using aniline as nucleophilic catalyst by the formation of an aniline Schiff base intermediate.^{[302],[303],[304]}

For coupling to the target proteins' amino groups, carboxylic group residues at the Biotin-Heparin or Biotin-Enoxaparin subunits were used upon activation with sNHS or NHS. As heparin is insoluble in non-aqueous buffers, the reaction was performed in PBS with the water-soluble 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide hydrochloride (EDC-HCl) (see Scheme 4.4).^{[307],[308],[309]} Each repeating disaccharide unit contains one free carboxylic group. If not indicated differently, the sNHS or NHS excess is chosen according to the average number of subunits in the heparin chain to fully activate the side groups.



Scheme 4.4.: Carboxylic group activation at heparin side chains.

4.2.2.2. Reactivity of Heparin-Based Biotinylation Reagents

Mass spectrometric analysis of the heparin modified model protein BSA is in most cases not possible. Heparin tags are very heavy, leading to extreme peak broadening and flattening in mass spectra. Furthermore, the reagent mass is distributed over a wide mass range and modified proteins exhibit different properties during sample preparation, possibly resulting in losses of the polar molecule during desalting on an unpolar material. If BSA is modified with higher reagent excesses, the effects are increased by the addition of multiple tags to a single protein molecule. Activation of the carboxylic side-groups with high ratios of sNHS can additionally lead to protein crosslinking and further increase the mass of the resulting complex. Only the detection of modification with LMW heparins like enoxaparin in a 1:1 ratio to BSA and low sNHS excess was possible (see Figure 4.10).

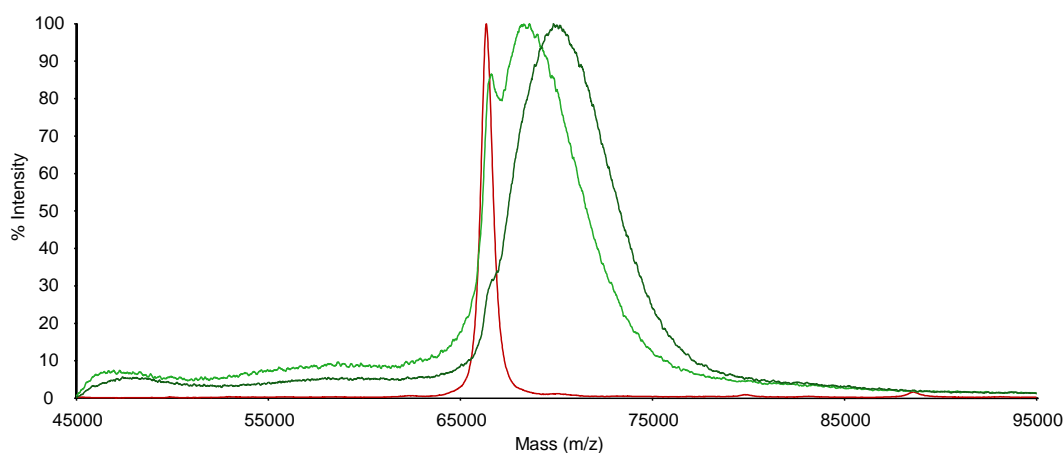


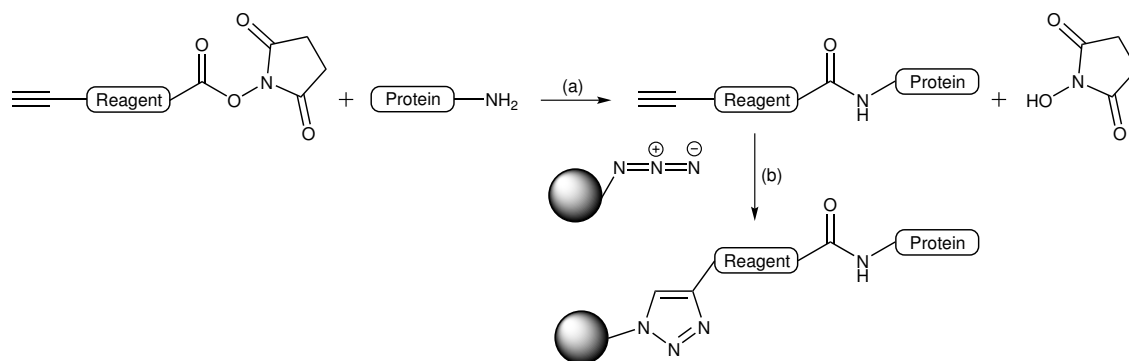
Figure 4.10.: Biotinylation of model protein BSA with Enoxaparin-sNHS in a ratio of 1:1. Mass spectra of unmodified BSA (red), 3x (light green) and 5x (dark green) activated Enoxaparin-sNHS are shown.

4.2.3. Biotin-Free System

4.2.3.1. Click Chemistry-Based Approach

Click chemistry is a general term describing a way to join subunits by a heteroatom linkage in a modular and stereospecific way, with high atom economy, under simple reaction conditions, with high yields and with no or inoffensive by-products.^[310] One of the most popular reactions is the Huisgen 1,3-dipolar cycloaddition of azides and terminal alkynes, that is catalyzed by Cu(I) using either a Cu(I) halide with an amine base and a stabilizing ligand or a stable Cu(II) salt under reductive conditions as Cu(I) source.^{[311],[312],[313]} The system is suitable for bioorthogonal reactions, as azides and alkynes are not present in biomolecules and can therefore be used for selective modification. As the reaction can proceed under mild conditions and in aqueous buffers under physiological pH and temperature, the so-called click reaction has been applied in many biological systems: Functionalisation of proteins, lipids, nucleic acids and glycans by chemical derivatisation or metabolic labelling and many applications including detection and imaging, enrichment and dynamic monitoring are reported.^{[314],[315],[316],[317],[318],[319],[320]}

The labelling of cell surface proteins with alkyne-tagged reagents, followed by click chemistry mediated capturing on azide-modified beads would provide a bioorthogonal enrichment platform (see Scheme 4.5). Avoidance of the biotin system could prevent the co-purification of intrinsically biotinylated, high-abundant, intracellular proteins co-enriched on streptavidin-sepharose.



Scheme 4.5.: Protein modification and click-chemistry-mediated capturing. (a) Modification of protein amino groups with NHS-ester derivatives. (b) Bioorthogonal capturing based on click reaction.

The usage of a Cu(I) source as catalyst is essential for the click reaction success, but might favor a certain degree of protein precipitation in the desired application.^{[321],[322]} Furthermore, the Copper-mediated click protocols using standard Cu(I) stabilizers such as TBTA in aqueous buffers exhibit slow reaction kinetics.^{[323],[324]} Moreover, the standard Cu(I) stabilizer TBTA also leads to protein precipitation, as evaluated in a BSA solution by adding TBTA to a final concentration of 4 mM.

To overcome the toxicity of copper during applications in living systems, a catalyst-free variant of the click reaction has been developed.^[325] The terminal alkyne is exchanged by a cyclooctyne: The acetylene bond massively deforms the bond angles leading to an enormous ring strain that favors the alkyne reaction.^[325] The method has been widely exploited in various applications such as metabolic labelling in cell systems for imaging purposes, cell surface glycan labelling for the study of internalisation kinetics as well as metabolic labelling in mice using azido glycans.^{[326],[327],[328]}

The usage of a cyclooctyne such as DBCO instead of a terminal alkyne might provide an easy way to capture alkyne-labelled proteins without dealing with copper-related effects on proteins. Still, the Cu-mediated click capturing is worth to be examined, as the generation of the terminal alkyne compound is cheaper and there are other Cu(I) stabilizers such as THPTA not causing protein precipitation.^[323]

4.2.3.2. Synthesis and Reactivity of Click Chemistry-Based Reagents

Two novel click chemistry-based reagents have been designed and synthesized, either with an terminal alkyne (SMCC-Cys-(L-Asp)₃-PA-alkyne) or a cyclooctyne (SMCC-Cys-(L-Asp)₃-DBCO-alkyne) (see Figure 4.11). The linker structure corresponds to the structure used for the peptide-based biotinylation reagents. Activation of the reagent has been performed as already described for the peptide-based biotinylation reagents (see chapter 4.2.1.1): A C-terminal cysteine residue was exploited for site-specific coupling to a bifunctional N-hydroxysuccinimide/maleimide crosslinker within 2.5 h at room temperature.

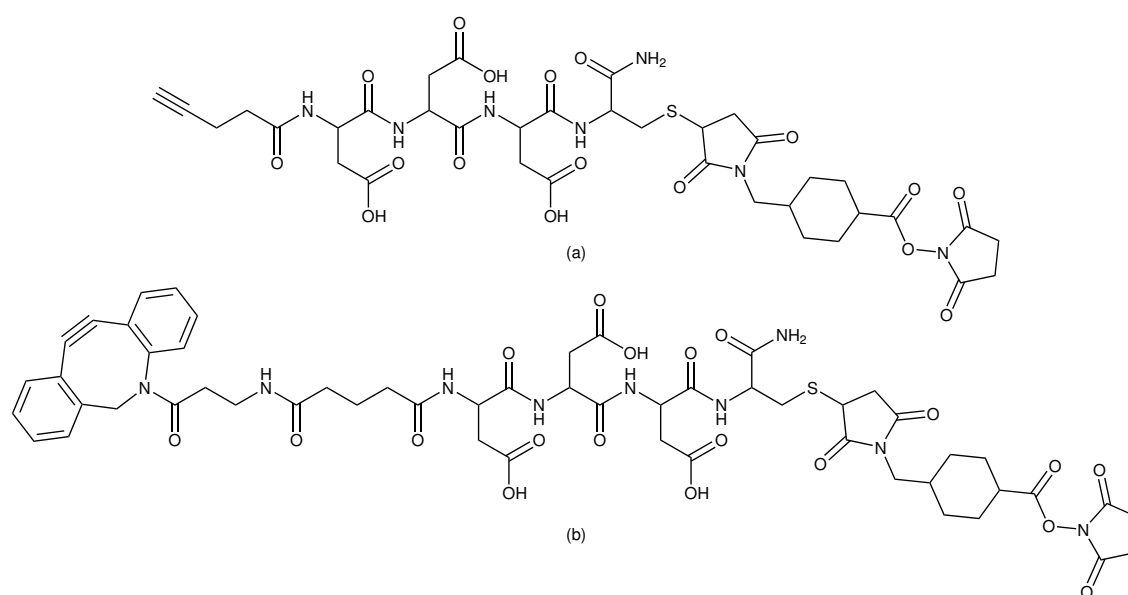


Figure 4.11.: Novel click-chemistry-based reagents. (a) SMCC-Cys-(L-Asp)₃-PA-alkyne, MW 879.85 g/mol; (b) SMCC-Cys-(L-Asp)₃-DBCO-alkyne, MW 1172.19 g/mol.

The reactivity was assessed with BSA as model protein as described before (see chapter 4.1.3). As expected from the structural similarity, the reactivity of SMCC-Cys-(L-Asp)₃-PA-alkyne and SMCC-Cys-(L-Asp)₃-DBCO-alkyne (see Figure 4.12) correspond to the reactivity of the peptide-based biotinylation reagents (compare Figure 4.7). As only minimal peptide amount was synthesized for initial tests, only a selection of labelling ratios were acquired for the latter reagent. Slight differences between both click chemistry-based reagents most likely occur due to minimally varying water content in the dried peptide influencing hydrolysis, an effect similar to batch-to-batch variability.

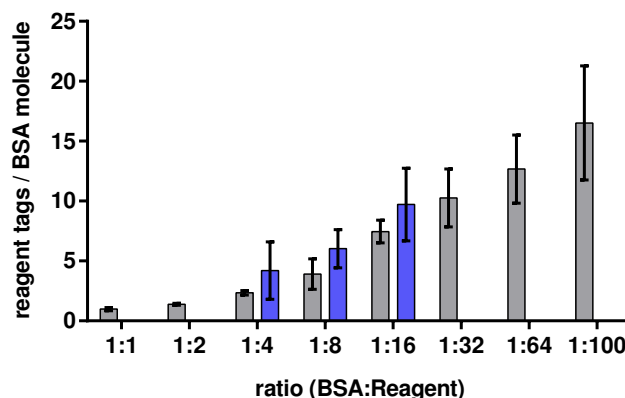


Figure 4.12.: Labelling of model protein BSA in different ratios with SMCC-Cys-(L-Asp)₃-PA-alkyne (grey) and SMCC-Cys-(L-Asp)₃-DBCO-alkyne (blue), analysis with linear MALDI-MS (mean and standard deviation, six to eight data points per condition).

In vitro labelling of cells with the novel click chemistry-based reagents SMCC-Cys-(L-Asp)₃-PA-alkyne and SMCC-Cys-(L-Asp)₃-DBCO-alkyne with subsequent capturing on azide-modified beads is currently set up.

4.3. *In Vitro* Validation of Novel Biotinylation Reagents

4.3.1. Workflow *In Vitro* Validation

A proof-of-principle study of the novel biotinylation reagents was performed in comparison to the commercial reagents Sulfo-NHS-LC-biotin and NHS-PEG₁₂-biotin. *In vitro* biotinylation was examined with HeLa as model cell line (see Figure 4.13). Biotinylation efficacy on cells was visualised by FACS analysis, confocal laser scanning microscopy (CLSM) or Western Blot using streptavidin conjugates for detection. Mass spectrometric analysis of the biotinylated proteome fraction was performed as comparative analysis of cells biotinylated with Sulfo-NHS-LC-biotin, NHS-PEG₁₂-

biotin, SMCC-Cys-(L-Asp)₃-biotin, SM(PEG)₆-Cys-(L-Asp)₃-biotin, Biotin-Heparin-sNHS or Biotin-Enoxaparin-sNHS and PBS-treated cells as negative control.

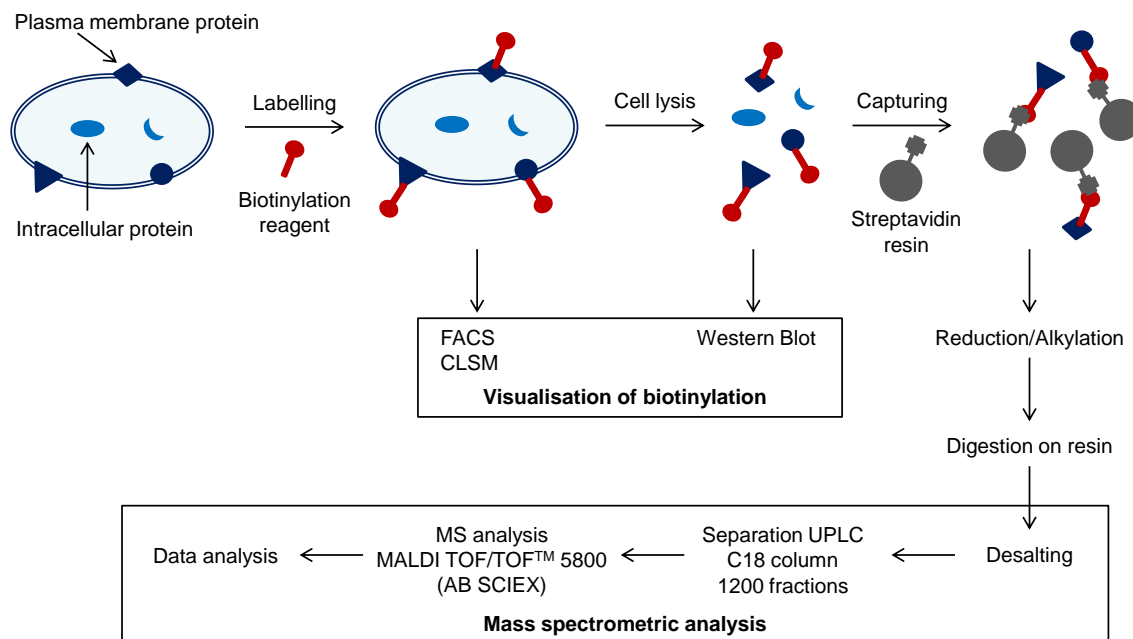


Figure 4.13.: Workflow *in vitro* validation.

4.3.2. Visualisation of Cell Surface Biotinylation

4.3.2.1. Peptide-Based Biotinylation Reagents

The degree of cell surface biotinylation with different amounts of biotinylation reagents could be assessed by FACS analysis of counted cell fractions after non-tryptic detachment with 10 mM EDTA, biotinylation and staining with a streptavidin Pacific Blue conjugate. HeLa cells were incubated with different amounts of biotinylation reagent, covering a range from 1:0.01 to 1:20. The molecular ratios refer to a median protein amount of 300 µg (6 nmol) protein content per 1×10^6 cells with a median protein weight of 50 kDa.^[130] The FACS reactivity profile corresponds to the results obtained by reactivity tests on BSA (see chapter 4.2.1.2): Sulfo-NHS-LC-biotin is the most reactive biotinylation reagent, reaching signal saturation at a ratio of 1:5 (30 nmol reagent per 1×10^6 cells) with three log-shifts difference to the control. The non-specifically activated NHS-β-Ala-(L-Asp)₃-biotin is less reactive compared to the site-specifically activated SMCC-Cys-(L-Asp)₃-biotin, proven by smaller peak shifts at the same ratios as well as a saturation level at two vs. three log-shifts, i.e. demonstrating the saturation of the cell surface with less and more spacious reagent tags (see Figure 4.14). The reactivity of both novel peptide-based reagents SMCC-Cys-(L-Asp)₃-biotin and SM(PEG)₆-Cys-(L-Asp)₃-biotin is comparable as shown in Figure 4.15.

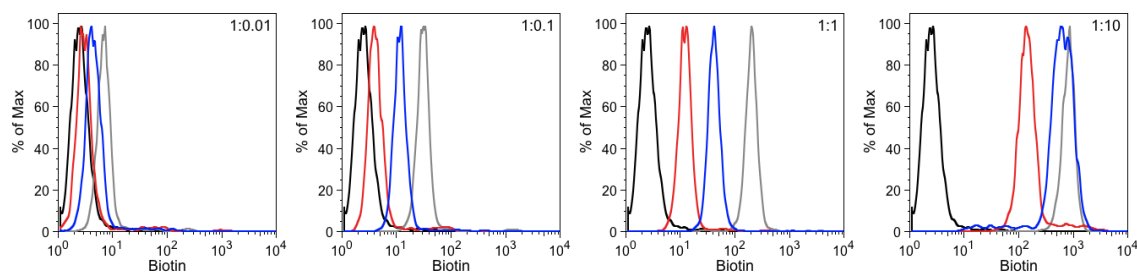


Figure 4.14.: FACS analysis of HeLa cells biotinylated with NHS- β -Ala-(L-Asp)₃-biotin (red) published by Strassberger *et al.*^[12] in comparison to Sulfo-NHS-LC-biotin (grey) and SMCC-Cys-(L-Asp)₃-biotin (blue). Biotinylation ratios between 0.01 eq (0.06 nmol) and 10 eq (60 nmol) reagent per 1×10^6 cells. Non-biotinylated control depicted in black.

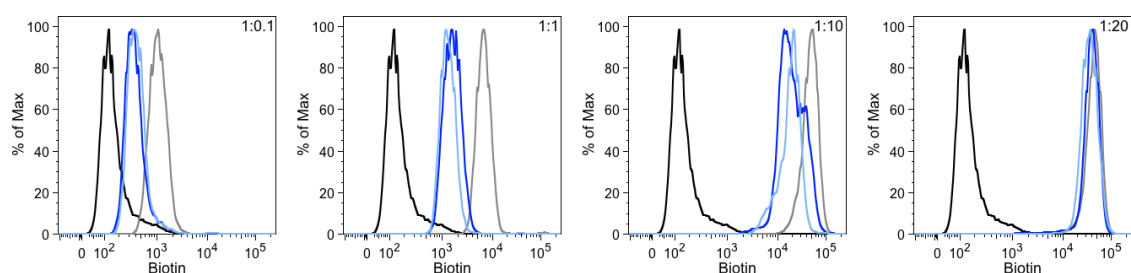


Figure 4.15.: FACS analysis of HeLa cells biotinylated with the novel peptide-based reagents SMCC-Cys-(L-Asp)₃-biotin (dark blue) and SM(PEG)₆-Cys-(L-Asp)₃-biotin (light blue) in comparison to commercial Sulfo-NHS-LC-biotin (grey). Biotinylation ratios between 0.1 eq (0.6 nmol) and 20 eq (120 nmol) reagent per 1×10^6 cells. Non-biotinylated control depicted in black.

Biotinylation of plasma membrane proteins could be shown by CLSM analysis (see Figure 4.16). Membrane staining with a streptavidin Alexa 488 conjugate could be observed with all reagents, but the total signal intensity was lower with the peptide-based reagents compared to the commercial ones. These findings were in line with the previously examined differences in reactivity. Therefore, laser properties were optimized for both cases individually.

4.3.2.2. Heparin-Based Biotinylation Reagents

The reactivity of heparin-based reagents on cells could be assessed by FACS analysis of counted cell fractions after detachment, biotinylation and staining with a streptavidin Pacific Blue conjugate. Heparin-based biotinylation reagents only reach a maximum of one log-shift for fully biotinylated cell surfaces, directly corresponding to the enormous size of the reagent tags allowing less biotin reagent molecules per surface area compared to the commercial or peptide-based reagents. Additional to the incubation with different amounts of the biotinylation reagent, differential

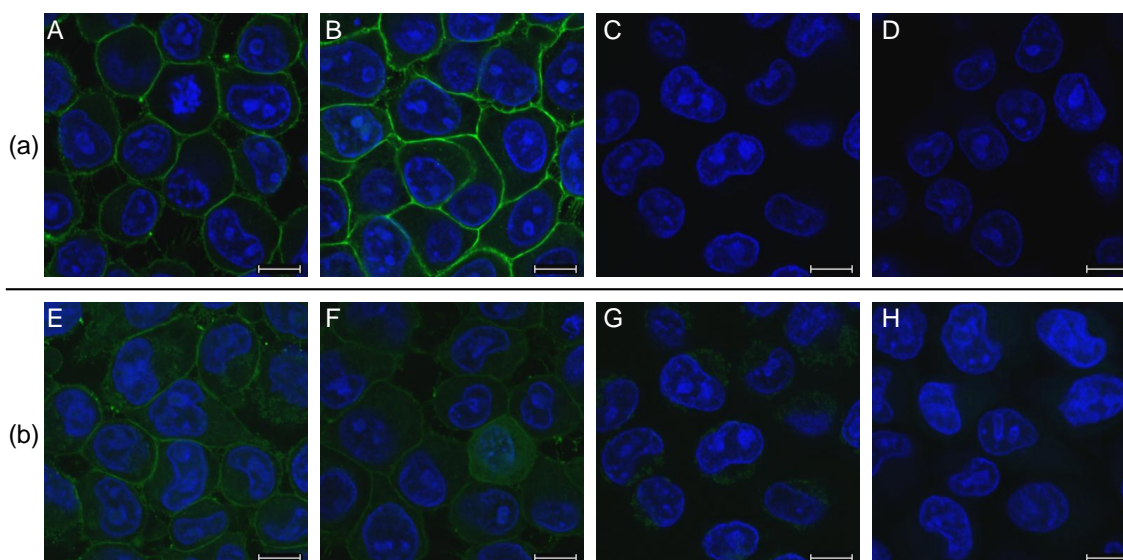


Figure 4.16.: CLSM of biotinylated HeLa cells. DAPI (blue), biotin staining (green), scale bars: 10 μ m. Commercial reagents (a) and peptide-based reagents (b) were acquired with different laser intensities to depict membrane biotinylation. Biotinylation with Sulfo-NHS-LC-biotin A, NHS-PEG₁₂-biotin B, SMCC-Cys-(L-Asp)₃-biotin E or SM(PEG)₆-Cys-(L-Asp)₃-biotin F, corresponding PBS treated controls (C,G) and staining controls for Sulfo-NHS-LC-biotin and SMCC-Cys-(L-Asp)₃-biotin without biotin staining (D,H).

grades of NHS-activation at the heparin chains were examined (see Figure 4.17). No unspecific binding of non-activated heparin-based biotinylation reagents to the cells was observed. Very low NHS-activation degrees led to reagent tags with decreased binding potential. Maximum log-shift was examined by full sNHS-activation of every disaccharide subunit.

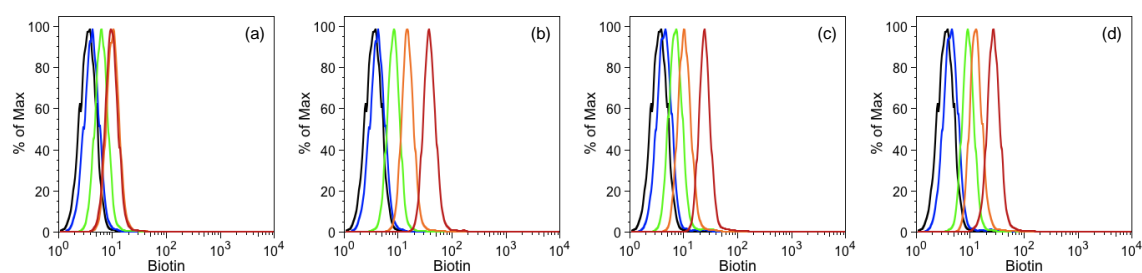


Figure 4.17.: FACS analysis of HeLa cells biotinylated with the novel heparin-based reagents Biotin-Heparin-sNHS and Biotin-Enoxaparin-sNHS. Biotinylation reagent excess 1 eq (green), 10 eq (orange) and 20 eq reagent (red). Biotinylation with 2x (a) and 5x (b) activated Biotin-Enoxaparin-sNHS or with 5x (c) and 25x (d) activated Biotin-Heparin-sNHS. Non-biotinylated control depicted in black, treatment with non-activated Biotin-Enoxaparin (a,b) or Biotin-Heparin (c,d) in blue.

4.3.3. Sample Preparation for Mass Spectrometric Analysis

For subsequent mass spectrometric analysis, HeLa cells were treated with biotinylation reagents while still adherent on the cell culture flask to avoid an increase in dead and therefore membrane permeable cells by the detachment procedure. 5 μmol biotinylation reagent per cell culture flask containing about 5×10^6 cells have been used. To perform a comparable analysis, the same amount was used for every reagent. The optimal amount for biotinylation of adherent cells was determined by FACS analysis. As the reagents on adherent cells in the cell culture flask are about ten-fold diluted compared to biotinylation of detached cell aliquots as used in chapter 4.3.2, the total reagent amount had to be increased. Maximal log-shift was achieved for all reagents using 1 μmol biotinylation reagent per 1×10^6 adherent cells (see Figure 4.18).

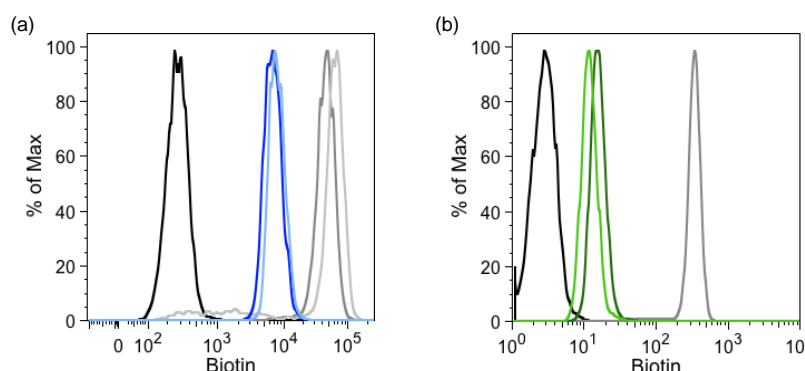


Figure 4.18.: FACS analysis of HeLa cells biotinylated on the cell culture flask. (a) Biotinylation with novel peptide-based reagents SMCC-Cys-(L-Asp)₃-biotin (dark blue) and SM(PEG)₆-Cys-(L-Asp)₃-biotin (light blue) in comparison to commercial Sulfo-NHS-LC-biotin (dark grey) and NHS-PEG₁₂-biotin (light grey). (b) Biotinylation with novel heparin-based reagents Biotin-Heparin-sNHS (dark green) and Biotin-Enoxaparin-sNHS (light green) in comparison to commercial Sulfo-NHS-LC-biotin (grey). Non-biotinylated control depicted in black in (a,b).

The cell viability is of greatest importance for a valuable mass spectrometric dataset of the cell surface proteome. During cell death, the plasma membrane undergoes substantial changes including increased membrane permeability, which would lead to labelling of intracellular proteins with the biotinylation reagents.^{[329],[330],[331]} Viability was assessed under different conditions with the Trypan Blue method. Non-tryptic detachment with 10 mM EDTA is slow, but necessary to preserve the surface proteins, however cell viability after subsequent biotinylation with Sulfo-NHS-LC-biotin is decreased to 71-79%, leading to an extreme contamination with labelled intracellular proteins. This is why for subsequent mass spectrometric analysis, cells were pre-washed with PBS and directly incubated with the biotinylation reagents in PBS while still attached to the cell culture flask. Reaction time was kept short to minimize negative effects on the cells and a subsequent quenching step inactivated non-reacted reagent before continuing with the detachment.

The novel peptide-based reagents were synthesized in dry DMSO to decrease hydrolysis rate of the activated carboxylic group to a minimum, and were diluted with PBS directly before use. DMSO is reported as membrane permeability enhancer increasing the cellular uptake as well as the solubility of small molecule drugs at low concentrations of 10-20% DMSO. Furthermore, a concentration dependent increase in cell death is observed, whereby the toxic concentration is cell line dependent.^{[332],[333],[334]} No cytotoxicity of DMSO on HeLa cells could be observed after 5 min incubation in up to 5% DMSO. Synthesis of novel peptide-based reagents was performed in highly concentrated solutions to ensure a final DMSO concentration on the cells below 1%.

Biotinylated proteins were captured out of full cell lysates on streptavidin-sepharose, starting with the same amount of total protein per sample for comparative analysis. Disulfide bonds were reduced and alkylated prior to tryptic digestion of the captured proteins on the streptavidin beads to enhance the enzymes's cleavage efficiency. Capture efficiency could be demonstrated by Western Blot analysis of HeLa cells biotinylated with commercial Sulfo-NHS-LC-biotin (see Figure 4.19). Due to decreased reactivity of the peptide- and heparin-based reagents, Western Blot analysis of the novel reagents is not possible without loading vast protein amounts severely compromising the quality of the analysis. FACS and CLSM analysis are the method of choice to reliably evaluate the biotinylation of cells.

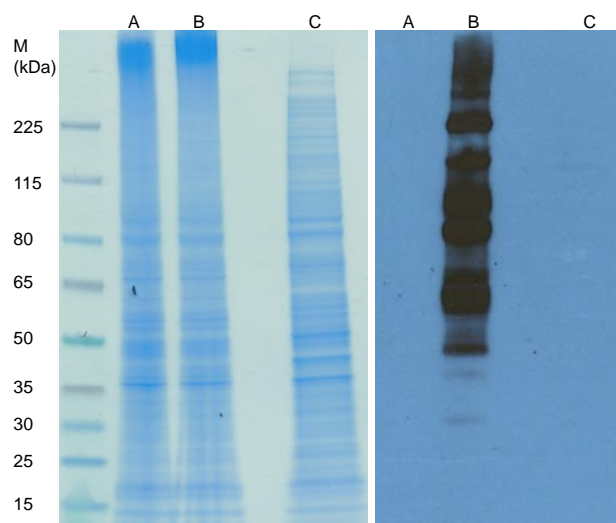


Figure 4.19.: SDS-PAGE and Western Blot analysis of cell lysate from HeLa. Cell surface protein biotinylation with Sulfo-NHS-LC-biotin (B), PBS treated cell lysate as negative control (A), supernatant of B after capturing with streptavidin-sepharose (C); 15 μ g total protein loaded per lane.

The captured and digested amount of biotinylated proteins resulting from originally 500 μ g of total protein was analysed per LC/MS run. According to experience with the LC/MS system and 2D peptide map densities, the biotinylated proteome fraction is estimated to be 1-2% of the total protein content within the cell lysate.

4.3.4. *In Vitro* Validation: Mass Spectrometric Analysis

4.3.4.1. Peptide-Based Biotinylation Reagents

Adherent HeLa cells were biotinylated with equal amounts of 5 μmol biotinylation reagent (SMCC-Cys-(L-Asp)₃-biotin, SM(PEG)₆-Cys-(L-Asp)₃-biotin, Sulfo-NHS-LC-biotin or NHS-PEG₁₂-biotin and PBS as negative control) per cell culture flask. For comparative analysis, capturing of biotinylated proteins was performed with identical total protein amounts of 500 μg . The analysis of technical triplicates per reagent was performed on a MALDI TOF/TOFTM 5800 system (AB SCIEX).

MS1 spectra were extracted using the Data Explorer[®] Software. For visualisation of successful chromatographic separation and differences in captured protein amount, 2D peptide maps were generated (see Figure 4.20).

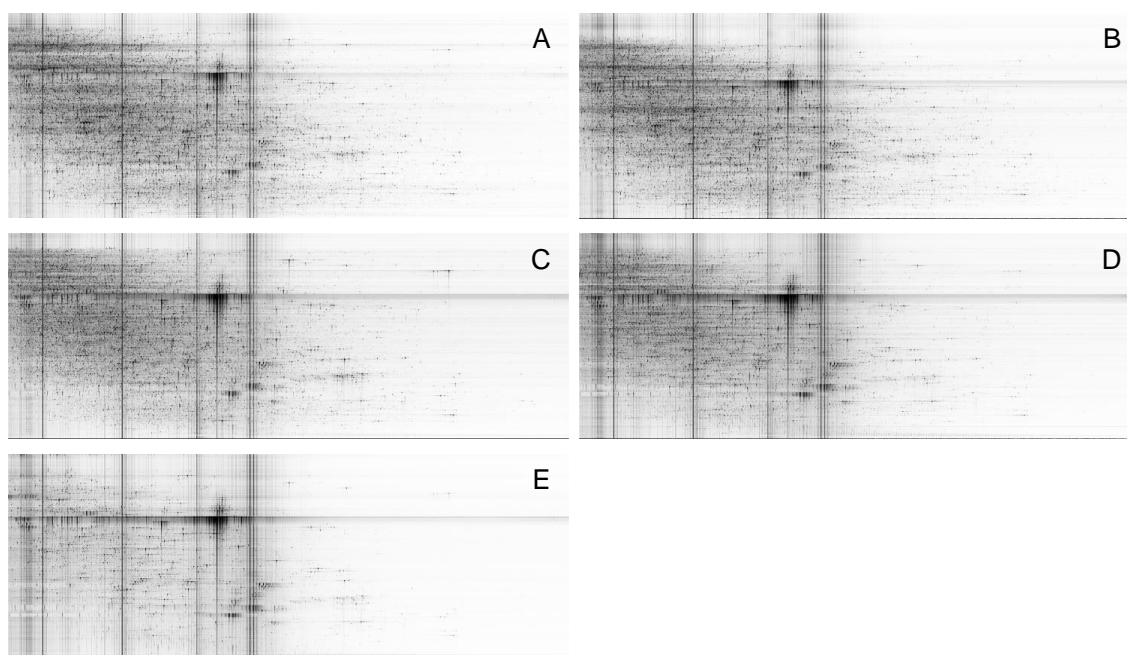


Figure 4.20.: 2D peptide maps of *in vitro* biotinylated samples (exemplary maps). Biotinylation with Sulfo-NHS-LC-biotin (A), NHS-PEG₁₂-biotin (B), SMCC-Cys-(L-Asp)₃-biotin (C), SM(PEG)₆-Cys-(L-Asp)₃-biotin (D), PBS (negative control) (E). m/z ratio (x-axis), UPLC fractions (y-axis) and normalized signal intensity as grey scale. Normalization to internal standard at m/z 1411 with standard peak height set to 1000, peak height capping at 500.

MS/MS spectra were searched against a human database using the ProteinPilotTM Software with the ParagonTM algorithm.^[335] Peptide proteotypicity was determined against a human database without isoforms with the in-house developed PepSir software.^[282] Degenerated peptide entries as well as peptides not reaching the 95% confidence limit filter were not included for further analysis.

The summaries of proteotypic peptides were used to analyse the differences in identified peptides/proteins with the particular biotinylation reagents. The informa-

tion obtained from triplicate measurements was summed for analysis to compensate for technical run-to-run differences. In total, 1574 proteins were identified, thereof 1090 proteins with more than one proteotypic peptide in at least one reagent group (termed without one hit wonders, in brackets) (see Suppl. Table B.3). 1374 (1015) proteins were identified with commercial Sulfo-NHS-LC-biotin, 983 (830) with NHS-PEG₁₂-biotin. With the novel peptide-based reagent SMCC-Cys-(L-Asp)₃-biotin 1096 (888) proteins were obtained, 517 (488) with SM(PEG)₆-Cys-(L-Asp)₃-biotin. Numbers and overlap between samples are shown in Figure 4.21.

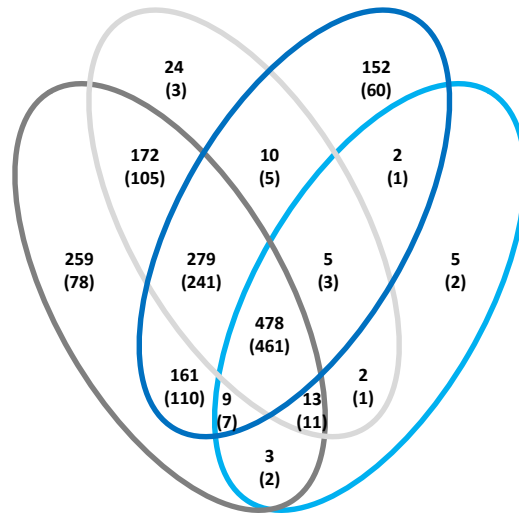


Figure 4.21.: Venn diagram showing numbers of identified proteins (without one hit wonders) of *in vitro* biotinylated samples. Commercial reagents Sulfo-NHS-LC-biotin (dark grey), NHS-PEG₁₂-biotin (light grey) and peptide-based reagents SMCC-Cys-(L-Asp)₃-biotin (dark blue), SM(PEG)₆-Cys-(L-Asp)₃-biotin (light blue).

Within the PBS-treated negative control 65 (62) proteins were identified. The degree of non-specific binding of highly abundant proteins to the streptavidin-sepharose is low. The top hits identified with high peptide numbers within the negative control (see Table 4.1) are not binding non-specifically to the streptavidin resin: Biotin is a covalently bound cofactor for carboxylases, leading to their co-purification with the biotinylated target proteins.^{[336],[337]}

ID	Protein name	Gene name	PBS	LC	PEG	SMCC	SMPEG
Q13085	Acetyl-CoA carboxylase 1	ACACA	99	101	91	98	104
P11498	Pyruvate carboxylase	PC	76	57	58	69	64
P05165	Propionyl-CoA carboxylase alpha chain	PCCA	29	21	15	29	30
Q96RQ3	Methylcrotonyl-CoA carboxylase subunit alpha	MCCC1	21	13	11	26	21
O00763	Acetyl-CoA carboxylase 2	ACACB	8	8	4	6	13

Table 4.1.: Top hits identified in PBS control. Numbers of identified proteotypic peptides are shown for the negative control (PBS), the commercial reagents Sulfo-NHS-LC-biotin (LC) and NHS-PEG₁₂-biotin (PEG) as well as for both peptide-based reagents SMCC-Cys-(L-Asp)₃-biotin (SMCC) and SM(PEG)₆-Cys-(L-Asp)₃-biotin (SMPEG).

The identification of a high number of proteotypic peptides from low-abundant plasma membrane proteins is the prerequisite for a stable quantification in biomarker research. The enrichment of low-abundant plasma membrane proteins and the simultaneous decrease in identification of high-abundant intracellular proteins is exemplarily demonstrated in the following by comparison to a full proteome dataset of the HeLa cell line obtained by Nagaraj *et al.*^[130] The summary of all identified peptides within the literature dataset was filtered for proteotypicity using the in-house developed PepSir software with the same human database used for the surface proteome dataset. Nagaraj *et al.* identified 10255 proteins, thereof 10% annotated to the plasma membrane. Based on MS1 signal intensities, Nagaraj *et al.* estimated cellular abundance (copies per cell) of the identified proteins by intensity based absolute quantification (iBAQ): The sum of all peptide peak intensities divided by the number of all theoretically observable peptides was translated to an absolute protein amount based on the total protein amount in the analyzed sample.^{[130],[42]} Hereby, a median copy number of 18 000 molecules per protein and cell was obtained; the lower abundant half of the proteome only accounted for 2% of its mass.^[130]

The significantly decreased identification of high-abundant intracellular proteins is demonstrated exemplarily with the 20S core unit of the 26S proteasome (see Table 4.2). The ubiquitin-proteasome pathway is a major pathway for protein degradation within cells.^[338] The eukaryotic 26S proteasome consists of a 20S core and two 19S regulatory subunits at the ends.^[339] The barrel-shaped 20S core unit is composed of 28 subunits, namely two copies of seven α - and seven β -type subunits forming outer and inner rings, the PSMB8 subunit can hereby replace PSMB5.^{[339],[340]} The

ID	Protein name	Gene name	copies per cell	full proteome	LC	PEG	SMCC	SMPEG	PBS
P25786	Proteasome subunit alpha type-1	PSMA1	1.2×10^6	9	-	-	-	-	-
P25787	Proteasome subunit alpha type-2	PSMA2	2.2×10^6	13	-	-	-	-	-
P25788	Proteasome subunit alpha type-3	PSMA3	1.3×10^6	16	-	-	-	-	-
P25789	Proteasome subunit alpha type-4	PSMA4	1.2×10^6	15	1	-	1	-	-
P28066	Proteasome subunit alpha type-5	PSMA5	3.3×10^6	21	-	-	-	-	-
P60900	Proteasome subunit alpha type-6	PSMA6	1.8×10^6	21	2	3	1	-	-
O14818	Proteasome subunit alpha type-7	PSMA7	2.8×10^6	14	-	-	-	-	-
P20618	Proteasome subunit beta type-1	PSMB1	2.8×10^6	12	-	-	-	-	-
P49721	Proteasome subunit beta type-2	PSMB2	2.7×10^6	3	1	-	1	-	-
P49720	Proteasome subunit beta type-3	PSMB3	1.3×10^6	4	-	-	-	-	-
P28070	Proteasome subunit beta type-4	PSMB4	1.7×10^6	3	-	-	-	-	-
P28074	Proteasome subunit beta type-5	PSMB5	1.0×10^6	4	1	-	1	-	-
P28072	Proteasome subunit beta type-6	PSMB6	1.3×10^6	7	-	-	-	-	-
Q99436	Proteasome subunit beta type-7	PSMB7	4.3×10^5	6	-	-	-	-	-
P28062	Proteasome subunit beta type-8	PSMB8	1.7×10^5	8	-	-	-	-	-

Table 4.2.: Subunits of the 20S core of the 26S proteasome identified in the *in vitro* dataset. Copy numbers per cell as well as the full proteome data taken from Nagaraj *et al.*^[130] Numbers of identified proteotypic peptides are shown for the full proteome dataset, the commercial reagents Sulfo-NHS-LC-biotin (LC) and NHS-PEG₁₂-biotin (PEG) as well as for both peptide-based reagents SMCC-Cys-(L-Asp)₃-biotin (SMCC) and SM(PEG)₆-Cys-(L-Asp)₃-biotin (SMPEG) and the negative control (PBS).

proteasome subunits are high-abundant within the cell (on average 1.7×10^6 copy numbers per subunit and cell). Whereas the identification was stable within the full proteome dataset of Nagaraj *et al.*, only a few subunits are identified with very low peptide numbers within the surface proteome dataset in all biotinylation reagents groups.

On the other hand, low copy number plasma membrane proteins were compared between the full proteome from Nagaraj *et al.* and the surface proteome dataset (see Table 4.3). Low-abundant proteins are under-represented in the full proteome dataset. Some proteins are not identified at all within the full proteome dataset, but also for identified, low-abundant proteins, the number of identified, proteotypic peptides is mostly small, impairing stable quantification of the protein. The numbers of identified proteotypic peptides is heavily increased across the surface proteome dataset with all different biotinylation reagents enabling a stable quantification.

ID	Protein name	Gene name	copies per cell	full proteome	LC	PEG	SMCC	SMPEG	PBS
Q9NPH3	Interleukin-1 receptor accessory protein	IL1RAP	200	-	14	13	10	12	-
Q14517	Protocadherin Fat 1	FAT1	260	10	40	35	50	37	-
Q02413	Desmoglein-1	DSG1	730	1	3	-	9	2	-
O14786	Neuropilin-1	NRP1	1300	1	14	13	14	11	-
Q6UVK1	Chondroitin sulfate proteoglycan 4	CSPG4	1500	3	68	70	67	60	-
P02751	Fibronectin	FN1	2000	9	10	7	16	18	-
Q6YHK3	CD109 antigen	CD109	3500	8	43	38	36	27	-
P23229	Integrin alpha-6	ITGA6	1.1×10^4	13	21	21	17	14	-
P26006	Integrin alpha-3	ITGA3	3.3×10^4	14	16	14	8	4	-
P32004	Neural cell adhesion molecule L1	L1CAM	5.4×10^4	18	49	48	55	48	-
Q14126	Desmoglein-2	DSG2	6.3×10^4	23	47	45	40	37	-
P13987	CD59 glycoprotein	CD59	5.4×10^5	8	7	6	4	3	-
Q4KMG0	Cell adhesion molecule-related/ down-regulated by oncogenes	CDON	-	-	2	1	2	2	-
Q8N441	Fibroblast growth factor receptor-like 1	FGFRL1	-	-	5	5	3	2	-

Table 4.3.: Plasma membrane proteins (examples) identified in the *in vitro* dataset. Copy numbers per cell as well as the full proteome data taken from Nagaraj *et al.*^[130] Numbers of identified proteotypic peptides are shown for the full proteome dataset, the commercial reagents Sulfo-NHS-LC-biotin (LC) and NHS-PEG₁₂-biotin (PEG) as well as for both peptide-based reagents SMCC-Cys-(L-Asp)₃-biotin (SMCC) and SM(PEG)₆-Cys-(L-Asp)₃-biotin (SMPEG) and the negative control (PBS).

Enrichment for cell surface proteins with all biotinylation reagents could be demonstrated by gene ontology (GO) analysis using the WEB-based GENE SeT AnaLysis Toolkit (Webgestalt).^{[283],[284]} *Cell surface* and *cell periphery* are the main enriched branches in the directed acyclic graph. Moreover, *cytosolic ribosome* is an enriched category: Within the dataset, ribosomal subunits are identified throughout all biotinylation reagent groups with a median number of three proteotypic peptides per protein. Ribosomal subunits belong to the highest abundant proteins in the HeLa cell line as shown by Nagaraj *et al.*: More than 70% of the identified ribosomal background proteins belong to the top 150 (< 2%) proteins in terms of abundance and have a median copy number of 4.9×10^6 molecules per cell.^[130] Reasons for the identification of intracellular background proteins will be elucidated later.

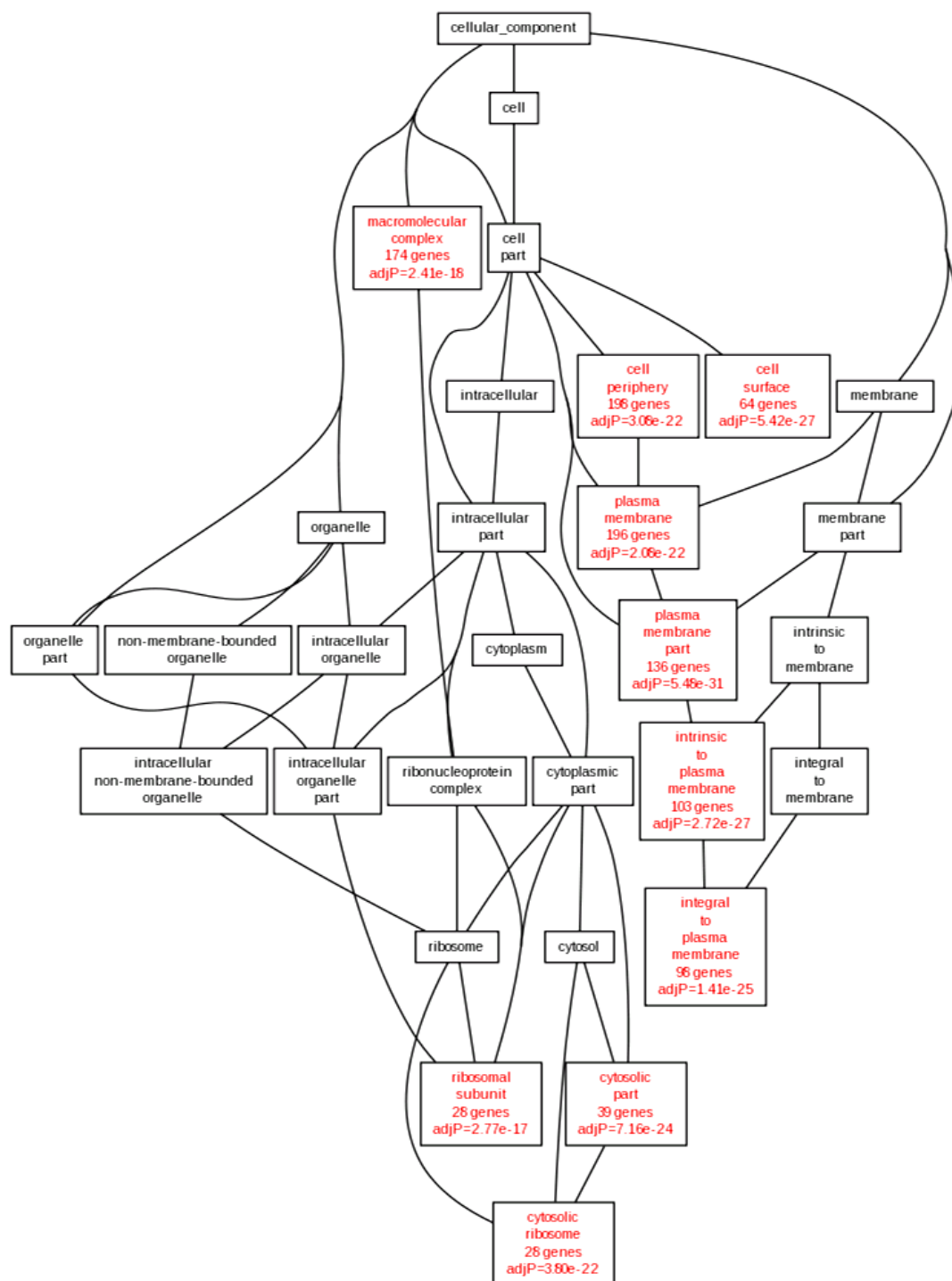


Figure 4.22.: Enrichment analysis of the cellular component of the core proteome (419 proteins) identified with all reagents, but not in the negative control. Visualisation as directed acyclic graph using Webgestalt.^{[283],[284]} Enriched GO categories are depicted in red, name of the GO category, number of included genes and adjusted p-value are shown.

To analyse differences in protein localization across the various biotinylation reagents, subcellular localization information was extracted from the UniProt website. The localization information was curated by screening for the categories plasma membrane, extracellular and secreted, membrane associated, intracellular and no information available (see Figure 4.23). Membrane associated proteins (depicted

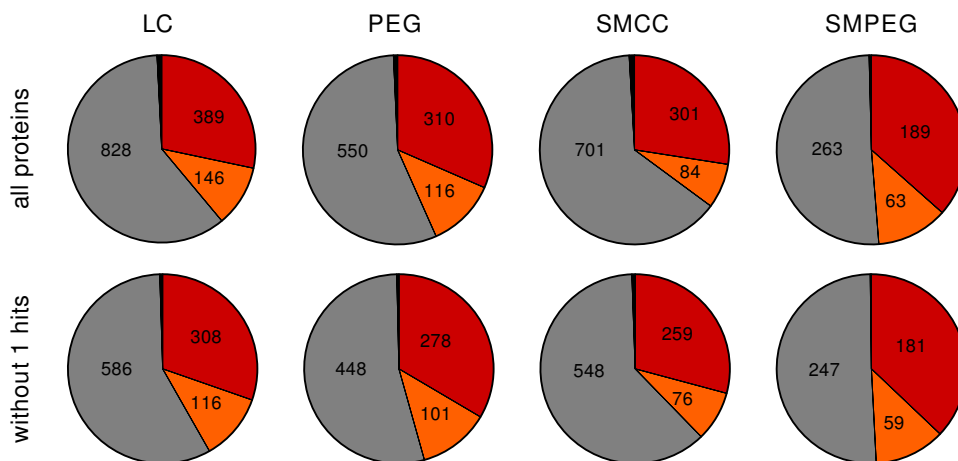


Figure 4.23.: Protein localization in the *in vitro* dataset. Percentage and total protein numbers are depicted. Categories: Plasma membrane, extracellular and secreted (red), membrane associated (orange), intracellular (grey), no annotation information available (black). Data (all proteins and without one hit wonders) for the commercial reagents Sulfo-NHS-LC-biotin (LC) and NHS-PEG₁₂-biotin (PEG) as well as for both peptide-based reagents SMCC-Cys-(L-Asp)₃-biotin (SMCC) and SM(PEG)₆-Cys-(L-Asp)₃-biotin (SMPEG) is shown.

in orange) contain hydrophobic transmembrane domains and are often reported to have receptor or cell adhesion functions. As plasma membrane proteome enrichment was used, those proteins are most likely also localised in the plasma membrane. The fraction of plasma membrane, secreted or membrane associated proteins is slightly increased, when proteins identified with only one proteotypic peptide are excluded from the analysis (without one hit wonders). The reason therefore is most likely the less stable identification of intracellular proteins, which is also visible in terms of numbers of identified proteotypic peptides per protein: Intracellular proteins are identified with an average number of 4.28 proteotypic peptides, whereas proteins belonging to the categories plasma membrane, extracellular or membrane associated are identified with 6.75 proteotypic peptides on average.

Plasma membrane, extracellular and membrane-associated proteins constitute between 38% (SMCC-Cys-(L-Asp)₃-biotin) and 49% (SM(PEG)₆-Cys-(L-Asp)₃-biotin) of all identified proteins (without one hit wonders). Both reagents with a pegylated linker (NHS-PEG₁₂-biotin, SM(PEG)₆-Cys-(L-Asp)₃-biotin) show a decreased reactivity: Whereas the fraction of surface proteins is increased, the total number of identified proteins is decreased. Size of the reagents and steric hindrance on the cell

surface may play a role. Moreover, also the previously determined lower reactivity of SMCC-Cys-(L-Asp)₃-biotin compared to Sulfo-NHS-LC-biotin is visible in terms of numbers of identified proteins. 38% (48% without one hit wonders) of the surface proteins are identified with all four biotinylation reagents. 16% (7%) are identified exclusively with Sulfo-NHS-LC-biotin, 9% (6%) only with SMCC-Cys-(L-Asp)₃-biotin, at which *protein binding* and *transmembrane transporter activity* are the main functional categories. The reactivity differences between the reagents is mirrored in the dataset in protein numbers, a fact also having an impact on the localization analysis. It was shown in chapter 4.2.1.2 that SMCC-Cys-(L-Asp)₃-biotin exhibits 59% reactivity (on average) compared to Sulfo-NHS-LC-biotin. In contrast, speaking of identified protein numbers, the number of identified surface annotated proteins is 72% (79% without one hit wonders) of the Sulfo-NHS-LC-biotin protein numbers. Both selectivity of the reagent and reactivity play contrary roles. The reactivity differences will be further examined later in this chapter.

Despite the surface proteome enrichment, a considerable fraction of intracellular proteins is identified within the dataset. Non-specific attachment of proteins to the streptavidin resin is negligible, as there are only few proteins identified in the PBS control. One reason is the sample preparation: The presence of some apoptotic cells with permeable plasma membranes leads to biotinylation of intracellular proteins. Especially very high-abundant proteins such as ribosomal subunits (see Figure 4.22) are therefore stably identified across all different reagents sets but the PBS control. Furthermore, co-purification of intracellular proteins attached to the labelled plasma membrane proteins is possible. A certain extent of plasma membrane crossing of small biotinylation reagents such as Sulfo-NHS-LC-biotin cannot be excluded: Biotin transport across the cell membrane is reported to be carrier-mediated at low concentrations, but non-saturated at high concentrations suggesting passive diffusion.^{[341],[342],[343]} An additional explanation can be found within the data analysis strategy: The analysis of subcellular localization relies on database entries of GO annotation, which might be incomplete or wrong. It is known, that antigens reported as intracellular can be translocated to the plasma membrane under specific conditions such as in a pre-apoptotic state or on malignant cells and cancer cell lines.^{[16],[17],[18],[19]}

In the following, similarity of technical replicates and reactivity differences between the biotinylation reagents is further examined. Relative quantification was performed using the in-house developed software MS_QBAT, a label-free method based on MS1 signal intensities of proteotypic peptides.^[282] The method is further described in chapter 3.8.4. Briefly, MS1 intensities were normalized against internal standard peptides, followed by annotation using the peptide summaries of proteotypic peptides. Individual features were extracted based on m/z, retention time and annotation. Data complexity was reduced by applying a S/N and feature length filter to obtain a more stable quantification. Hereby, 39% of the features could be dismissed on average, while 99.5% of the identifications were kept. On average, 458 712 peaks per biotinylated sample were extracted to 79 289 features and reduced to 50 669 features after the complexity reduction step.

Cross-alignments within technical triplicates were used to propagate annotation information, decreasing the variability from mass spectrometric performance differences and creating a more stable dataset (see Figure 4.24).

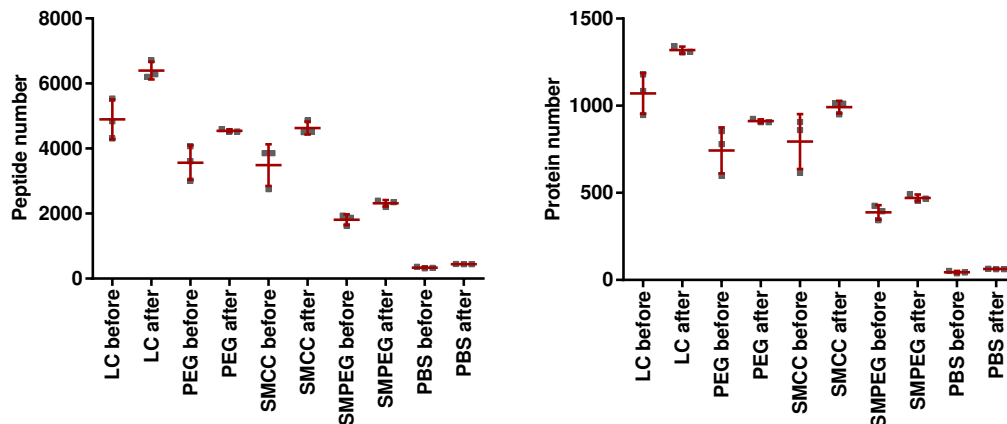


Figure 4.24.: Annotation propagation with MS_QBAT. Peptide/protein numbers before and after annotation propagation within technical triplicates (mean and standard deviation). Samples: Sulfo-NHS-LC-biotin (LC), NHS-PEG₁₂-biotin (PEG), SMCC-Cys-(L-Asp)₃-biotin (SMCC), SM(PEG)₆-Cys-(L-Asp)₃-biotin (SMPEG), negative control (PBS).

Pairs of technical replicates were compared by feature alignment followed by protein quantification. The variability between technical replicates is low as shown in Figure 4.25.

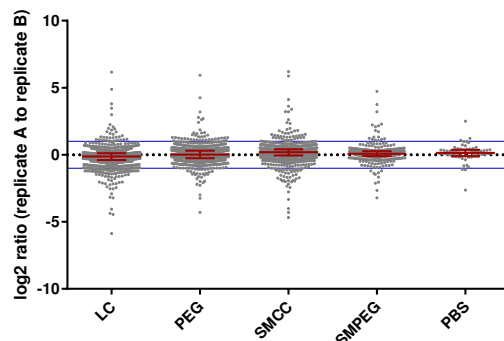


Figure 4.25.: Quantification of pairs of technical replicates, median with interquartile range. Blue dotted line ($y = 0$) indicates no regulation between samples, blue straight lines ($y = \pm 1$) illustrate two-fold up- or down-regulation. Commercial reagents Sulfo-NHS-LC-biotin (LC) and NHS-PEG₁₂-biotin (PEG), peptide-based reagents SMCC-Cys-(L-Asp)₃-biotin (SMCC) and SM(PEG)₆-Cys-(L-Asp)₃-biotin (SMPEG) and negative control (PBS).

Subsequently, technical triplicates were combined to super-samples via alignment by a genetic algorithm (GEAL alignment). For relative quantification, each super-sample was GEAL aligned against the Sulfo-NHS-LC-biotin super-sample (see

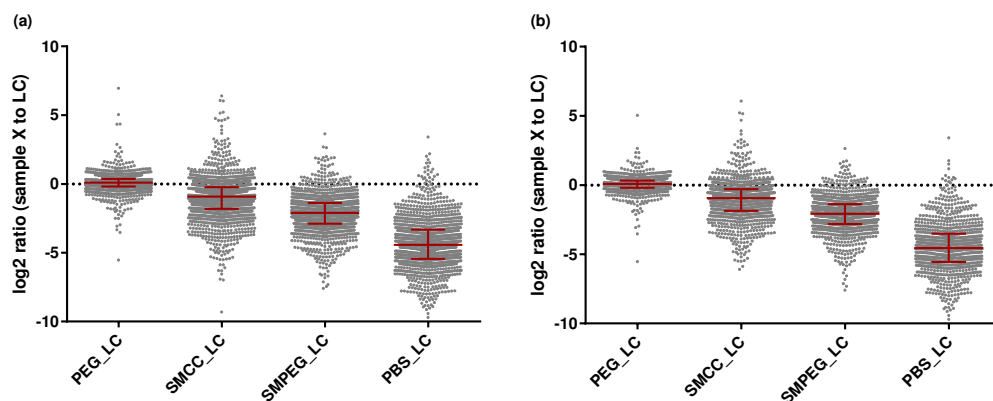


Figure 4.26.: Quantification of all biotinylation reagents (NHS-PEG₁₂-biotin (PEG), SMCC-Cys-(L-Asp)₃-biotin (SMCC), SM(PEG)₆-Cys-(L-Asp)₃-biotin (SMPEG) and negative control (PBS)) vs. Sulfo-NHS-LC-biotin, median with interquartile range. Blue dotted line ($y = 0$) indicates no regulation between samples. All proteins (a) or without one hit wonders (b).

Suppl. Table B.4 and Figure 4.26). In total, 1526 proteins could be quantified, thereof 1037 with more than one proteotypic peptide. Compared to the variability between technical replicates, the protein distribution is widespread, and narrowing a bit, if proteins quantified with only one proteotypic peptide are excluded from analysis as outliers are reduced. The variability between the commercial reagents Sulfo-NHS-LC-biotin and NHS-PEG₁₂-biotin is low. The peptide-based reagents exhibit a larger difference to commercial Sulfo-NHS-LC-biotin visualized by a widespread distribution. As already expected in terms of identified protein numbers, the decreased reactivity is evident by a shift of the median with a two-fold down-regulation with SMCC-Cys-(L-Asp)₃-biotin, and a five-fold down-regulation with SM(PEG)₆-Cys-(L-Asp)₃-biotin. Between the negative control and Sulfo-NHS-LC-biotin, 1274 (962) proteins are quantified, which is a sign for a presence vs. absence situation: To enable relative quantification, MS_QBAT replaces non-existing values with an artificially introduced background intensity value, resulting in a more than twenty-fold down-regulation (median).

The *in vitro* proof-of-principle study could demonstrate that the novel peptide-based reagents stably identify proteins annotated to the cell surface, whereas the identification of intracellular proteins is decreased. Previously determined reactivity differences across the biotinylation reagents were also mirrored in the mass spectrometric dataset.

4.3.4.2. Heparin-Based Biotinylation Reagents

Adherent HeLa cells were biotinylated with equal amounts of biotinylation reagent (Biotin-Heparin-sNHS, Biotin-Enoxaparin-sNHS, Sulfo-NHS-LC-biotin and PBS as negative control) with at least 1 μmol per 10^6 cells. For comparative analysis, capturing of biotinylated proteins was performed with identical total protein amounts. The

analysis of technical duplicates per reagent was performed on a MALDI TOF/TOFTM 5800 system (AB SCIEX). Data from consecutive mass spectrometric evaluation of different sample preparation strategies was acquired as singular measurements. Data extraction, database searches and extraction of proteotypic peptides were performed as described in chapter 4.3.4.1. To account for mass spectrometric performance differences and run-to-run variability, annotations were propagated across samples biotinylated with the same reagent based on MS1 data using MS_QBAT.

693 (598 without one hit wonders) proteins have been identified with Biotin-Enoxaparin-sNHS, 756 (619) proteins with Biotin-Heparin-sNHS, 1259 (928) proteins with Sulfo-NHS-LC-biotin and 37 (35) proteins in the negative control (see Suppl. Table B.5a). Overlap between technical duplicates was more than 95% in terms of protein identification. Subcellular localization was analysed using information from the UniProt website (see Figure 4.27). The surface proteome fraction was signifi-

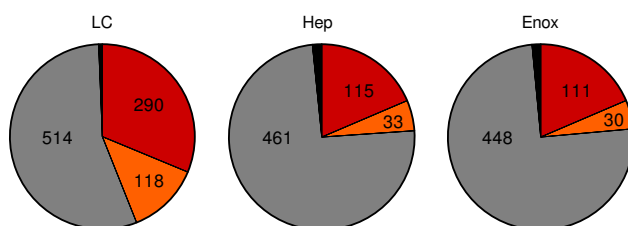


Figure 4.27.: Protein localization in the *in vitro* dataset. Percentage and total protein numbers are depicted, all proteins without one hit wonders. Categories: Plasma membrane, extracellular and secreted (red), membrane associated (orange), intracellular (grey), no annotation information available (black). Data for the commercial reagent Sulfo-NHS-LC-biotin (LC) as well as for both heparin-based reagents Biotin-Heparin-sNHS (Hep) and Biotin-Enoxaparin-sNHS (Enox) are shown.

cantly reduced with the heparin-based reagents compared to Sulfo-NHS-LC-biotin. To identify the reason for the reduced identification of the desired protein classes, an enrichment analysis was performed using Webgestalt based on all proteins identified with at least two peptides with both heparin-based reagents (see Figure 4.28).^{[283],[284]} As already shown in chapter 4.3.4.1, some high-abundant background proteins belonging to the *cytosolic ribosome* are enriched. For the heparin-based reagents, around half of all identified proteins belong to the highly enriched categories *nuclear part*, *heterocyclic compound binding/nucleotide binding/RNA binding*.

NHS-activation of the carboxylic group was shown to work efficiently and to be crucial for protein binding: The treatment of adherent cells with non-activated Biotin-Heparin led to the identification of only 21 (20) proteins, in accordance with the negative control (see Suppl. Table B.5c). In order to identify reasons for the co-purification of nuclear proteins mentioned above, different sample preparation strategies have been tested (see Suppl. Table B.5b). Variation of the number of charges by activation with nine- (half activated) or twenty-fold (fully activated heparin chains) excess of sNHS did not influence the result.

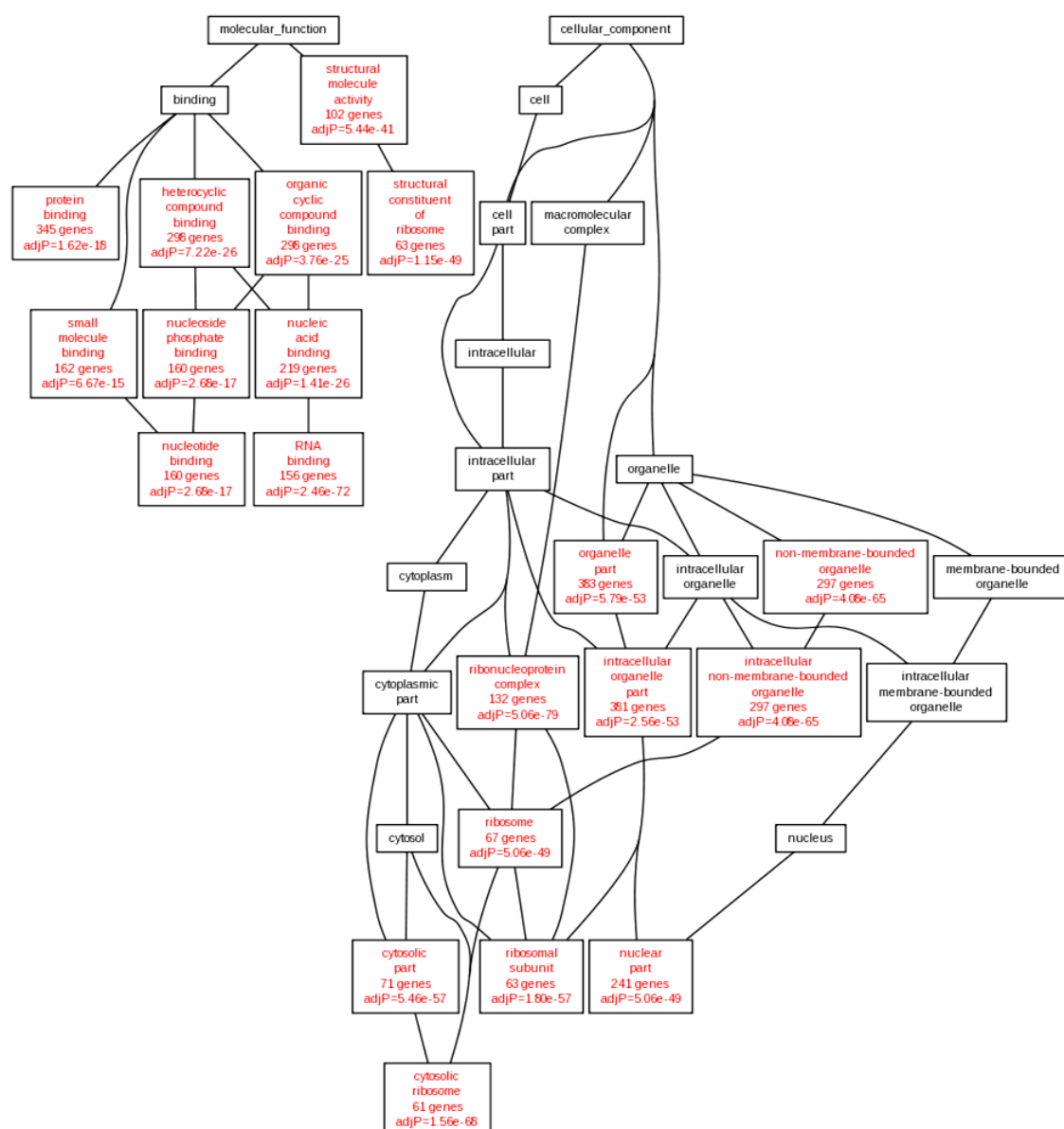


Figure 4.28.: Enrichment analysis of cellular component and molecular function of 545 proteins identified within all sample based on Biotin-Heparin-sNHS and Biotin-Enoxaparin-sNHS (without one hit wonders). Visualisation as directed acyclic graph using Webgestalt.^{[283],[284]} Enriched GO categories are depicted in red, name of the GO category, number of included genes and adjusted p-value are shown.

With the twenty-fold activated Biotin-Heparin-sNHS, 649 (558) proteins were identified with an intracellular fraction of 76.5%, with the nine-fold activated Biotin-Heparin-sNHS 639 (551) proteins were identified, thereof 76.6% with intracellular annotation (see Table 4.4). Moreover, captured protein samples have been subjected to various washing strategies. Gradients of 0-1 M sodium or potassium chloride, chang-

ing pH values as well as extraction in alkaline sodium carbonate solution are reported to disrupt non-covalent protein-protein or heparin-protein interactions.^{[344],[345],[346]} Therefore, washing with high-salt buffer (2 M NaCl), buffers with changing pH (pH 5, pH 8, pH 12) and carbonate extraction were examined on samples treated with Biotin-Heparin-sNHS (see Table 4.4). No significant change of the fraction of identified intracellular proteins was obtained with any strategy. Furthermore, the influence of delipidation with organic solvents (1-butanol/diisopropylether) was evaluated, a technique reported for efficient release and solubilisation of plasma membrane proteins from the surrounding lipids (see Table 4.4).^[347] Both with Biotin-Heparin-sNHS and the commercial reagent Sulfo-NHS-LC-biotin, the result was not changed, i.e. delipidation was shown to be unnecessary when working in an *in vitro* system. Reduction of disulfide bridges with TCEP and alkylation with IAA was used for all samples, as it is reported to improve tryptic digest, sequence coverage and the detection of cysteine-containing peptides in standard proteomic workflows.^{[348],[349]} Pre-incubation of the adherent cells with an excess of free DNA to non-covalently bind free nuclear proteins deriving from apoptotic cells did neither change the number of identified proteins nor the identified surface proteome fraction (437 proteins without one hit wonders, thereof 75.7% intracellular) (see Suppl. Table B.5c).

Sample	Preparation	Protein number	% intracellular
20x Biotin-Heparin-sNHS	reduction/alkylation	558	76.5%
20x Biotin-Heparin-sNHS	reduction/alkylation/high-salt	535	77.8%
20x Biotin-Heparin-sNHS	reduction/alkylation/pH	413	78.0%
20x Biotin-Heparin-sNHS	reduction/alkylation/carbonate	541	76.9%
20x Biotin-Heparin-sNHS	reduction/alkylation/delipidation	560	75.9%
9x Biotin-Heparin-sNHS	reduction/alkylation	551	76.6%
9x Biotin-Heparin-sNHS	reduction/alkylation/high-salt	554	76.5%
9x Biotin-Heparin-sNHS	reduction/alkylation/pH	435	77.9%
9x Biotin-Heparin-sNHS	reduction/alkylation/carbonate	545	76.1%
Sulfo-NHS-LC-biotin	reduction/alkylation	1095	61.5%
Sulfo-NHS-LC-biotin	reduction/alkylation/high-salt	1048	60.9%
Sulfo-NHS-LC-biotin	reduction/alkylation/pH	956	60.6%
Sulfo-NHS-LC-biotin	reduction/alkylation/carbonate	1084	60.9%
Sulfo-NHS-LC-biotin	reduction/alkylation/delipidation	1098	61.1%

Table 4.4.: Sample preparation tests on HeLa cells biotinylated with differentially activated Biotin-Heparin-sNHS or commercial Sulfo-NHS-LC-biotin. Numbers of identified proteins (without one hit wonders) and percentage of the intracellular fraction are shown. Sample preparation strategies are further described in the text.

Two hypotheses either based on non-covalent or on covalent binding can help to explain the co-enrichment of nuclear proteins with the heparin-based reagents. Reason for the severe enrichment of DNA- or RNA-binding proteins can be found in the structural similarity of the heparin chain to the DNA backbone. Since a long time, heparin beads have been widely used in affinity purification of heparin-

or DNA-binding proteins or other positively charged proteins such as transcription factors, nucleotide-binding enzymes, Antithrombin III, lipoprotein lipase or Interleukin-8.^{[350],[344],[351],[352],[353],[354]} The interaction between DNA and proteins are mostly based on ionic interactions, van der Waals contacts and H bonds, thereof about two thirds are reported to be non-specific with the sugar-phosphate backbone.^{[355],[356]} Charges of amino acid side chains are hereby distributed topologically so that the protein behaves like a dipole with the positive charge facing the DNA.^[357] The DNA blocking strategy might not show any significant effect, because the highly sulfated heparin chains mimic the structure of the negatively charged, ribose-based backbone of nucleotide chains and are reported to compete with the DNA for protein binding.^{[358],[359]}

Attachment of nuclear proteins to the heparin chains might not only occur during the labelling reaction due to a fraction of membrane permeable apoptic cells, but also within the cell lysates by non-covalent, polarity-based interactions. Due to the extremely high number of charged side groups in the heparin chains, the protein interaction might be too strong for effective elution of all non-covalently bound proteins when using the applied washing strategies. However, for elution of DNA-binding proteins from heparin columns, typically 0 M to 1 M, in rare cases also up to 2 M sodium or potassium chloride gradients are used.^{[350],[344],[360]} Therefore, the applied washing strategies, e.g. with 2 M sodium chloride, should at least decrease the fraction of nuclear proteins to support the hypothesis of a non-covalent interaction.

If covalent attachment is taken as a basis, the labelling reaction must occur on the adherent cells, before quenching of the reagent. Due to their size, heparin-based reagents are unlikely to cross the plasma membrane, therefore nuclear proteins must be derived from membrane-permeable apoptotic cells. The co-enrichment of the intracellular proteins cannot be dependent on pure protein abundance, as especially the nuclear fraction is severely enriched. Therefore, the labelling of those proteins might be mediated by a primary non-covalent interactions immediately becoming covalent due to the activated carboxylic groups. However, the reaction of activated carboxylic groups is so fast, that one can doubt the high specificity of the labelling reaction for a specific class of proteins. Nevertheless, compared to commercial Sulfo-NHS-LC-biotin, the same amount of heparin-based reagent has far more activated carboxylic groups (on average 25-fold activation) and therefore more possibilities for side-reactions.

The unavailability of further Biotin-Heparin unfortunately prevents the elucidation of this phenomenon. If possible, the coupling of non-activated Biotin-Heparin to streptavidin-sepharose, followed by addition of a whole cell lysate should be examined to shed some light on the undesired co-enrichment of nuclear proteins.

Importantly, the *in vitro* study of heparin-based biotinylation reagents has shown that the surface proteome enrichment is suffering from the ability of the heparin chains to interact with nucleic acid binding proteins leading to co-enrichment of nuclear proteins.

4.4. *In Vivo* Validation of Novel Biotinylation Reagents

4.4.1. Workflow *In Vivo* Validation

In vivo validation of the novel peptide- and heparin-based biotinylation reagents was performed via perfusion of healthy NSG mice in comparison to the commercial reagents Sulfo-NHS-LC-biotin and NHS-PEG₁₂-biotin (see Figure 4.29). Biotinylation efficacy was visualised by confocal laser scanning microscopy (CLSM) of tissue sections from kidney, liver, spleen, tongue and brain as well as by ELISA and Western Blot analysis using streptavidin conjugates for detection. Mass spectrometric analysis of the biotinylated proteome fraction was performed with kidney and liver samples from mice perfused with Sulfo-NHS-LC-biotin, NHS-PEG₁₂-biotin, SMCC-Cys-(L-Asp)₃-biotin, SM(PEG)₆-Cys-(L-Asp)₃-biotin, Biotin-Heparin-sNHS, Biotin-Enoxaparin-sNHS and PBS or non-perfused tissue as negative controls. Enrichment levels of selected protein candidates were further assessed via SRM analysis.

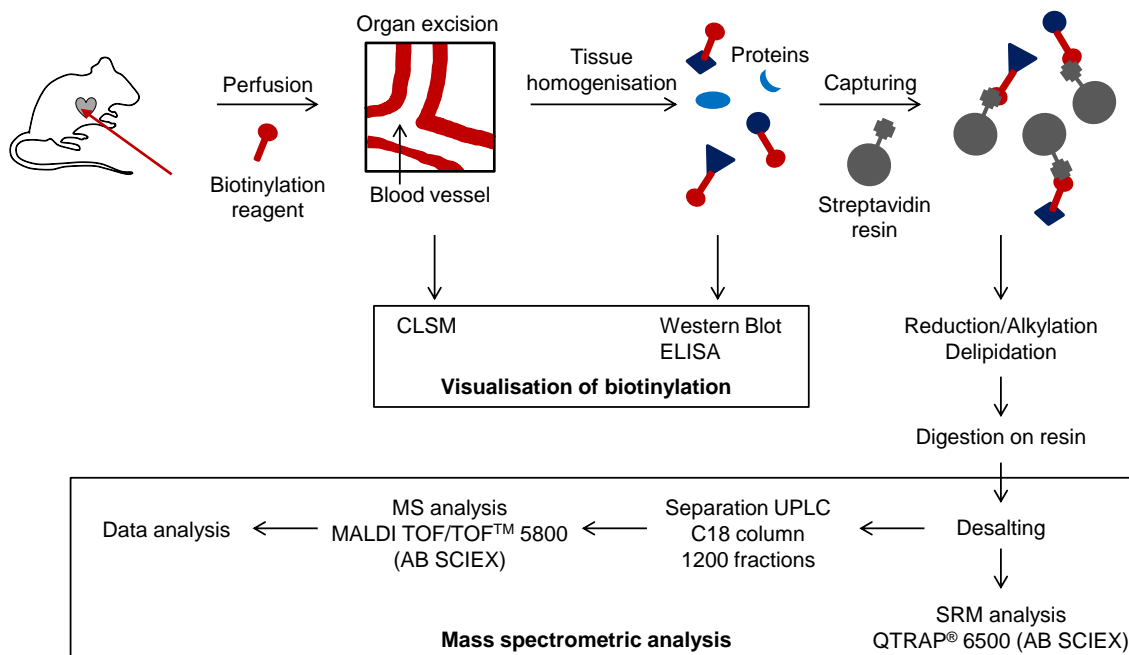


Figure 4.29.: Workflow *in vivo* validation.

4.4.2. Visualization of Biotinylation of Vascular Accessible Proteins

4.4.2.1. Biotinylation Depth Around Blood Vessels

For the development of novel diagnostics or targeted therapies, biomarkers exclusively expressed on the surface of diseased cells, in the surrounding extracellular matrix or on newly formed blood vessels have to be identified. Angiogenesis is a rare event in a healthy adult, occurring during wound healing and under pathological conditions.^[82] For instance, the formation of new blood vessels in tumour tissue is needed, as nutrient supply is essential for tumour growth and viability.^{[361],[84]} Hereby, tumour vessels exhibit an abnormal structure compared to the normal vasculature: Vessels are leaky and highly interconnected, endothelial cells have an aberrant morphology and are also distinct on molecular level.^{[79],[91],[89],[90]} Biodistribution studies proved targeted drugs to be stably enriched in the tumour tissue for several days.^{[362],[363],[364]} Other pathological situations with angiogenic events include chronic inflammatory diseases such as inflammatory bowel disease, rheumatoid arthritis, atherosclerosis, psoriasis or ocular angiogenesis in blinding eye disease such as proliferative diabetic retinopathy.^{[14],[94],[365],[102],[100]}

However, not all biomarkers within a diseased tissue are accessible by a specific monoclonal antibody or an antibody-drug conjugate: The distance around blood vessels reached by a diagnostic agent or a drug depends on the size and charge of the molecule (diffusion characteristics) as well as on the affinity to the target structure and the target concentration within the diseased tissue.^[72] Immunofluorescence studies revealed tumour penetration with various labelled antibodies against tumour-specific antigens: Clear vascular staining and an antibody-dependent penetration depth around blood vessels of at least three cell layers were detected.^{[366],[367]}

Stable identification and quantification of novel vascular-accessible targets by mass spectrometric analysis is enabled by vascular perfusion with biotinylation reagents labelling the surface of endothelial cells, cells in close proximity of the blood vessels and in the perivascular extracellular matrix.

In vivo perfusion of mice was performed according to the protocol from Rösli *et al.* with 25 μmol per biotinylation reagent and mouse.^[185] Immunofluorescent staining against CD31 as endothelial marker, nuclear staining with DAPI and biotin-staining with a streptavidin derivative was performed to reveal the biotin-labelled areas in various tissue samples. Successful vascular labelling and a biotinylation depth of a few cell layers around the blood vessels could be proven (see Figure 4.30) Kidney tissue exhibited the strongest signal, liver tissue was mainly stained around vascular structures. Furthermore, it could be shown that biotinylation reagents are able to enter highly vascularized regions such as the glomeruli in kidney tissue (see Figure 4.30A) or capillary structures between muscle fibers in the tongue (see Figure 4.30D).

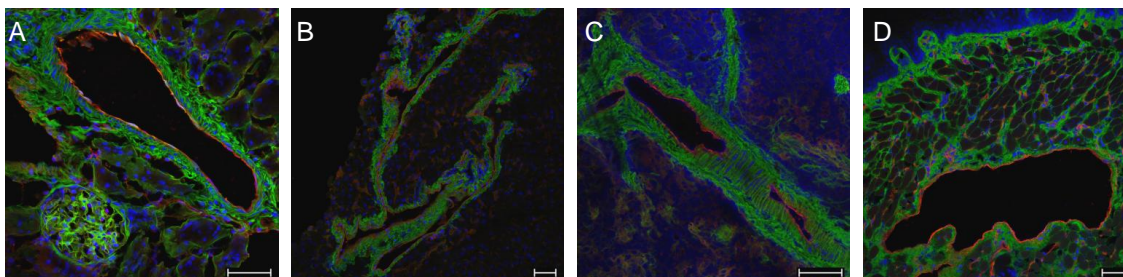


Figure 4.30.: Visualisation of biotinylation depth around blood vessels. DAPI (blue), biotin (green), endothelial marker CD31 (red), scale bars: 50 μm . Sections from mouse tissue perfused with SMCC-Cys-(L-Asp)₃-biotin: kidney A, liver B, spleen C and tongue D.

4.4.2.2. Biotinylation With Different Reagents

The successful biotinylation of kidney, liver, spleen, tongue and brain tissue was examined with commercial, peptide- (see Figure 4.31) and heparin-based reagents (see Figure 4.32). Vascular biotin labelling as well as the biotinylation depth around blood vessels are comparable across all reagents. Biotin staining from tissue sections of mice perfused with Biotin-Heparin-sNHS or Biotin-Enoxaparin-sNHS is weak, as expected for the enormous size of the reagents allowing less biotin derivatives per area compared to the commercial or peptide-based reagents. Therefore, laser properties were optimized for the two subgroups individually. PBS-perfused or non-perfused mice served as negative controls; corresponding staining controls to all samples are shown in Suppl. Figure A.1.

CLSM analysis of all samples taken for mass spectrometric validation of the novel peptide- and heparin-based biotinylation reagents as well as the corresponding staining controls are shown in Suppl. Figures A.2 and A.3 for kidney tissue and in Suppl. Figures A.5 and A.6 for liver tissue. Immunofluorescent staining intensity is both dependent on the biotinylation reagent and on the perfusion efficacy of a particular mouse. Biotin signal intensity was observed to be stronger for kidney tissue than for liver samples. Especially the signals for the heparin-based biotinylation reagents are close to the detection limit in liver tissue; non-specific and non-vascular background biotin-stain occurs in the corresponding PBS- or non-perfused negative controls.

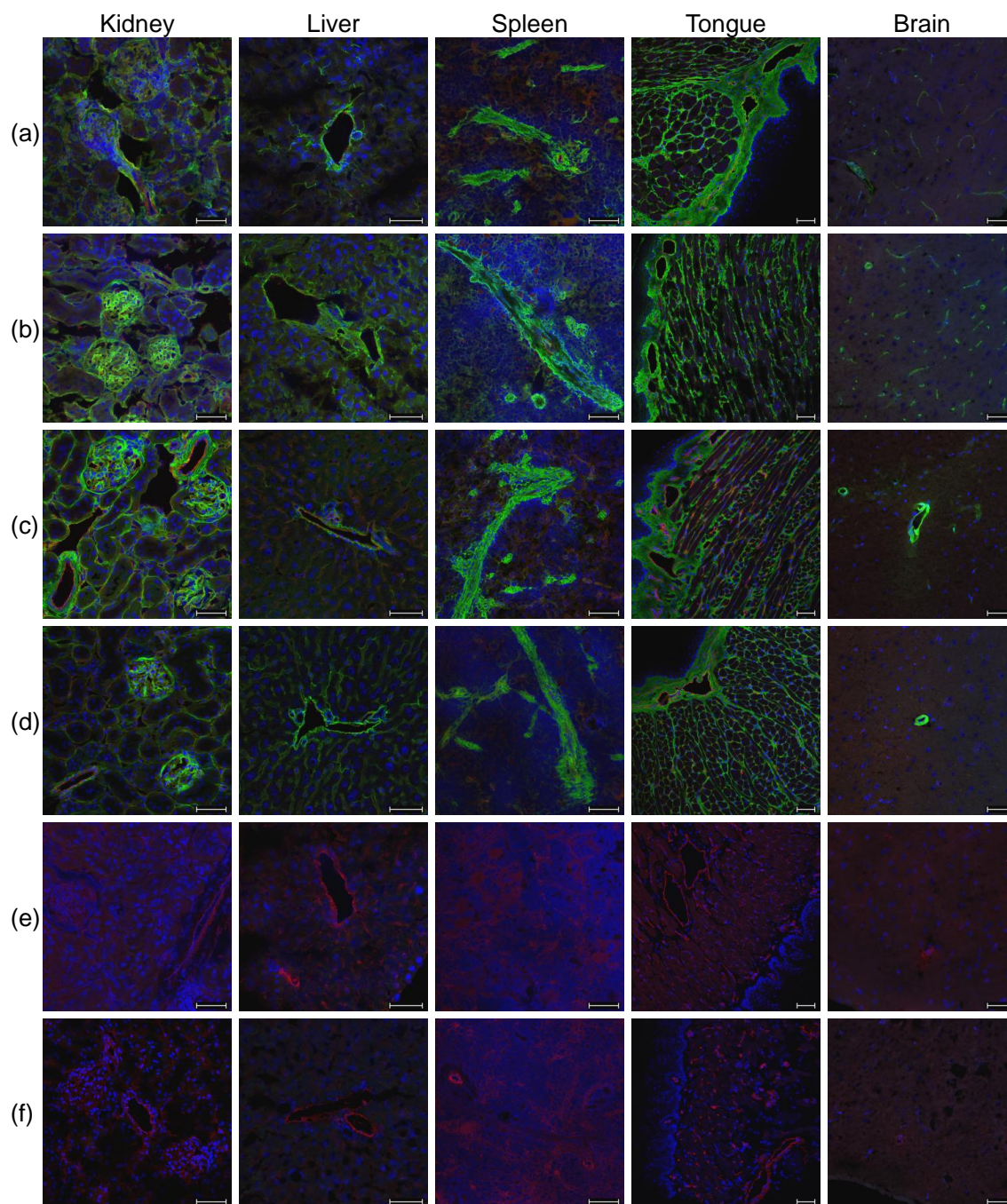


Figure 4.31.: CLSM of mouse tissue perfused with commercial and peptide-based reagents. DAPI (blue), biotin (green), endothelial marker CD31 (red), scale bars: 50 μ m. Sections from mouse tissue perfused with Sulfo-NHS-LC-biotin (a), NHS-PEG₁₂-biotin (b), SMCC-Cys-(L-Asp)₃-biotin (c), SM(PEG)₆-Cys-(L-Asp)₃-biotin (d), PBS perfused controls (e), non-perfused controls (f). Corresponding staining controls see Suppl. Figure A.1.

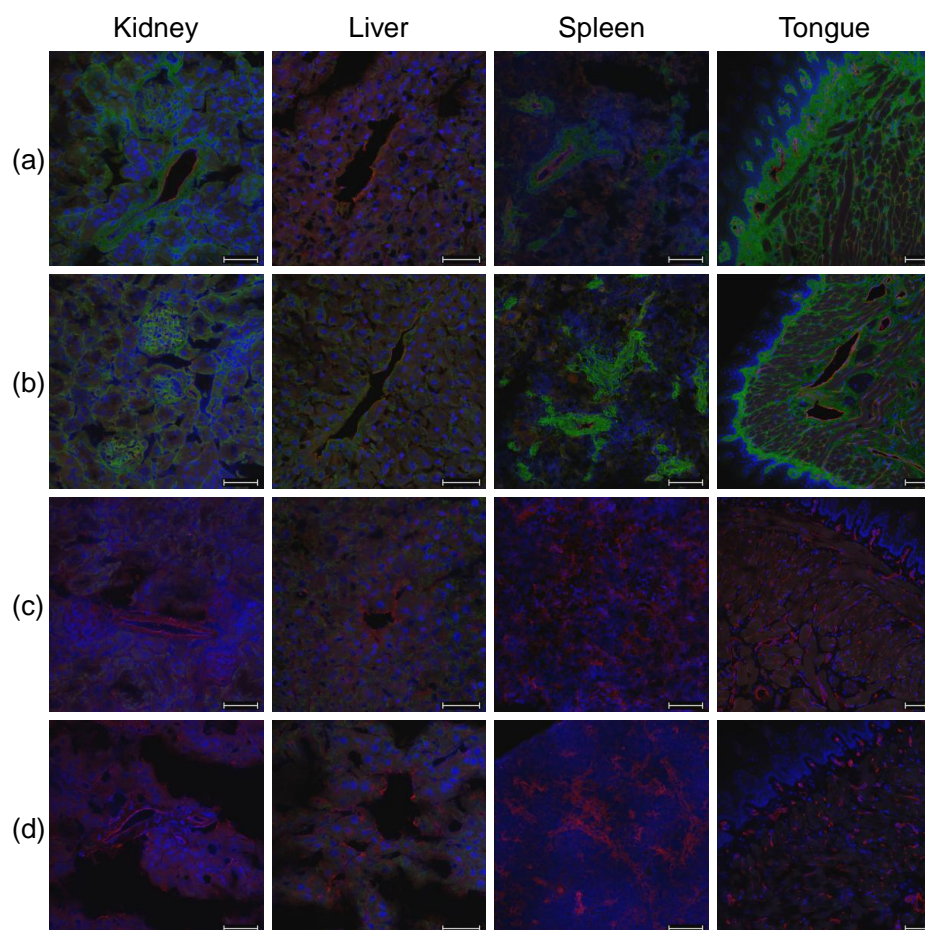


Figure 4.32.: CLSM of mouse tissue perfused with heparin-based reagents. DAPI (blue), biotin (green), endothelial marker CD31 (red), scale bars: 50 μm . Sections from mouse tissue perfused with Biotin-Heparin-sNHS (a), Biotin-Enoxaparin-sNHS (b), PBS perfused controls (c), non-perfused controls (d). Corresponding staining controls see Suppl. Figure A.1.

4.4.3. Sample Preparation for Mass Spectrometric Analysis

Healthy NSG mice were perfused with equal amounts of 25 μmol per biotinylation reagent (4 mice per reagent group, perfusion with Sulfo-NHS-LC-biotin, NHS-PEG₁₂-biotin, SMCC-Cys-(L-Asp)₃-biotin, SM(PEG)₆-Cys-(L-Asp)₃-biotin, Biotin-Heparin-sNHS, Biotin-Enoxaparin-sNHS, PBS- and non-perfused as negative controls), followed by perfusion with a Tris containing quenching solution to inactivate non-reacted reagent. Proteins were extracted from kidney and liver tissue by tissue homogenisation, sonication and one freeze-thaw cycle in a detergent-containing buffer. The amount of biotinylated proteins within the tissue homogenates was assessed by ELISA (see Figure 4.33) and Western Blot (see Figure 4.34).

All ELISA samples were measured against a BSA standard biotinylated with ten-fold excess of Sulfo-NHS-LC-biotin in a 1 to 10 ratio. Absolute ELISA biotinylation

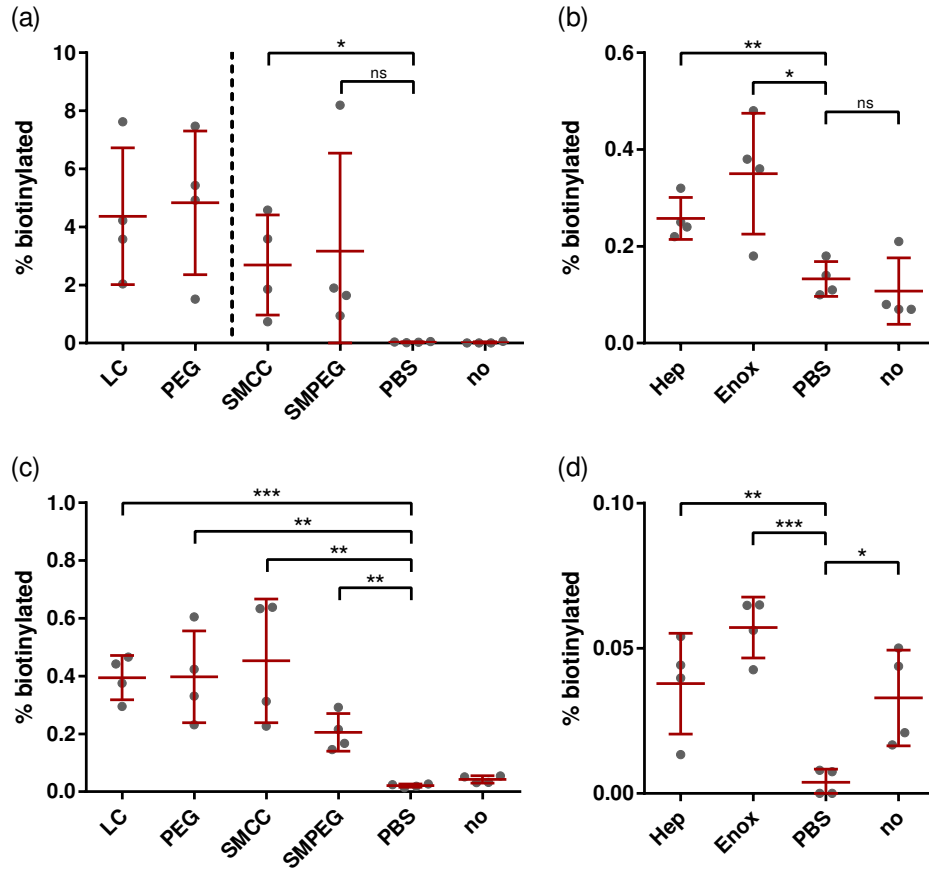


Figure 4.33.: ELISA of biotinylated mouse kidney (a,b) and liver (c,d) tissue. Biotinylation with Sulfo-NHS-LC-biotin (LC), NHS-PEG₁₂-biotin (PEG), SMCC-Cys-(L-Asp)₃-biotin (SMCC), SM(PEG)₆-Cys-(L-Asp)₃-biotin (SMPEG), Biotin-Heparin-sNHS (Hep), Biotin-Enoxaparin-sNHS (Enox), perfusion with PBS (PBS) or non-perfused (no) tissue as negative controls. Unpaired t-test with significance levels *** ($P \leq 0.0001$), ** ($P \leq 0.01$), * ($P \leq 0.05$), ns ($P > 0.05$).

values are therefore artificial and at best a rough estimation, but relative differences in biotinylation efficacy can be displayed within and across the sample groups. Differences in the biotinylation efficacy among biological replicates show the variation of perfusion efficacy, but also biological variation among non-perfused tissue from different mice. As already seen by CLSM measurements (see chapter 4.4.2.2), kidney tissue biotinylation is stronger than the biotinylation of liver tissue. Samples exhibit clear biotinylation within all reagent groups when compared to the PBS-perfused control. Nevertheless, reactivity differences between commercial, peptide- and heparin-based reagents are visible: For the highly biotinylated kidney tissue, commercial and peptide-based reagents cannot be evaluated within the same experiment, in order to ensure a linear correlation within the ELISA standard range. Heparin-based reagents exhibit low biotinylation levels. Background signal from PBS- or non-perfused mice is derived from endogenous biotin-containing proteins

(e.g. carboxylases), or, in the case of non-perfused mice, additionally from endogenous biotin serum-levels of about 4 ng/mL in blood.^[368]

Reactivity differences across the different reagents could also be shown by Western Blot analysis of kidney tissue homogenates using Streptavidin-HRP for detection. To obtain signals for liver tissue or heparin-based reagents, vast total protein amounts have to be loaded, severely restricting the quality of the Western Blot. The biotin-detection from perfusion with peptide-based reagents (see Figure 4.34) is only achievable following loading of triple total protein amount compared to the commercial reagents and by prolonging the Western Blot film exposure time.

Biotin-dependent carboxylases play pivotal roles in the carbohydrate, fatty acid and amino acid metabolism.^[369] The pyruvate carboxylase ($M = 127$ kDa) is localized in the mitochondrial matrix and is essential in kidney and liver tissue as main gluconeogenic organs.^[370] Methylcrotonoyl-CoA carboxylase ($M = 75$ kDa, subunit α) and Propionyl-CoA carboxylase ($M = 75$ kDa, α -chain) are also localized in the mitochondrial matrix and play pivotal roles in fatty acid and amino acid catabolism.^[369] The enzymes are known to be detectable by Western Blot analysis and are stably identifiable in the PBS- or non-perfused negative controls.^[370]

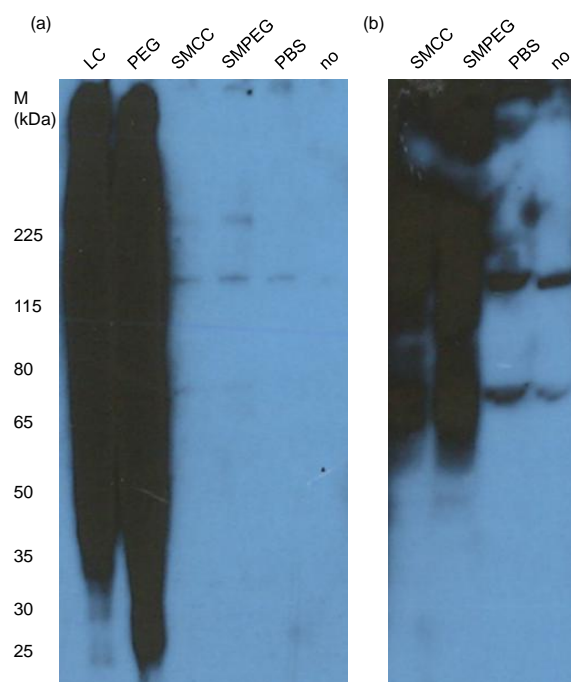


Figure 4.34.: Western Blots of biotinylated kidney homogenates. Perfusion with Sulfo-NHS-LC-biotin (LC), NHS-PEG₁₂-biotin (PEG), SMCC-Cys-(L-Asp)₃-biotin (SMCC), SM(PEG)₆-Cys-(L-Asp)₃-biotin (SMPEG), PBS (PBS) or non-perfused (no) as negative controls. 20 µg (a) or 60 µg (b) total protein loaded per lane. 5 min (a) or 20 min (b) exposure time.

4.4.4. Mass Spectrometric Analysis of Biotinylated Kidney Tissue

4.4.4.1. Discovery Experiment Kidney

For comparative proteomic analysis, capturing of biotinylated proteins was performed with equal amounts of 1.25 mg total kidney protein on 100 μ L streptavidin-sepharose. Samples were processed by reduction, alkylation, delipidation and tryptic digest. Peptides derived from 750 μ g total protein amount were examined per LC/MS run. Four biological replicates per reagent group (commercial reagents: Sulfo-NHS-LC-biotin, NHS-PEG₁₂-biotin, peptide-based reagents: SMCC-Cys-(L-Asp)₃-biotin, SM(PEG)₆-Cys-(L-Asp)₃-biotin, heparin-based reagents: Biotin-Heparin-sNHS, Biotin-Enoxaparin-sNHS, negative controls: PBS- or non-perfused) were analysed in two technical replicates each.

MS1 spectra were extracted using the Data Explorer[®] Software. Consistent chromatographic separation across the samples was visualised by 2D peptide mapping (see Suppl. Figure A.8). MS/MS spectra were searched against a mouse database using the ProteinPilot[™] Software with the Paragon[™] algorithm.^[335] Peptide proteotypicity was determined against a mouse database without isoforms using the in-house developed PepSir software.^[282] Degenerated peptide entries as well as peptides not reaching the 95% confidence limit filter were not included for further analysis. Peptide identification information was summed within one reagent group to compensate for technical run-to-run differences and to obtain a stable dataset.

In total, 1965 proteins were identified, thereof 1409 proteins with more than one proteotypic peptide in at least one reagent group (without one hit wonders, in brackets) (see Suppl. Table B.6). 1543 (1252) proteins were identified with commercial Sulfo-NHS-LC-biotin, 1142 (1026) with NHS-PEG₁₂-biotin. With the novel peptide-based reagent SMCC-Cys-(L-Asp)₃-biotin 1514 (1263) proteins were obtained, 1441 (1239) with SM(PEG)₆-Cys-(L-Asp)₃-biotin. 1209 (1053) proteins were found with Biotin-Heparin-sNHS, 1469 (1200) with Biotin-Enoxaparin-sNHS. 44% (58%) of all identified proteins within the datasets were found with every biotinylation reagent. Variability between heparin-based and commercial reagents is higher compared to peptide-based and commercial reagents: 1018 (957) proteins were identified with all commercial and peptide-based reagents, corresponding to a proportion of 56% (70%) within the subset, whereas 876 (831) proteins were identified with all commercial and heparin-based reagents, corresponding to a fraction of 48% (60%) within the subset. The overlap in protein identification between pairwise subsets of both commercial, peptide-based or heparin-based reagent is large with 72% (81%) for Sulfo-NHS-LC-biotin and NHS-PEG₁₂-biotin, 77% (85%) for Biotin-Heparin-sNHS and Biotin-Enoxaparin-sNHS and 77% (88%) for SMCC-Cys-(L-Asp)₃-biotin and SM(PEG)₆-Cys-(L-Asp)₃-biotin.

Within the PBS- or unperfused negative controls, 1008 (912) and 633 (597) proteins have been identified, respectively. As expected from the *in vitro* dataset, carboxylases carrying biotin as cofactor (e.g. Methylcrotonyl-CoA carboxylase, Pro-

pionyl-CoA carboxylase, Acetyl-CoA carboxylase) are within the top-hits, very stably identified with proteotypic peptide counts in the two-digit range. The identified background proteome is further elucidated in chapter 4.4.4.2.

Differences in protein localization across the various biotinylation reagents was analysed via extraction of subcellular localization information from the UniProt website. The localization information was curated by screening for the categories plasma membrane, extracellular and secreted, internal side of the plasma membrane, membrane associated, intracellular and no information available (see Figure 4.35). Similar to the *in vitro* dataset, plasma membrane, extracellular and membrane-

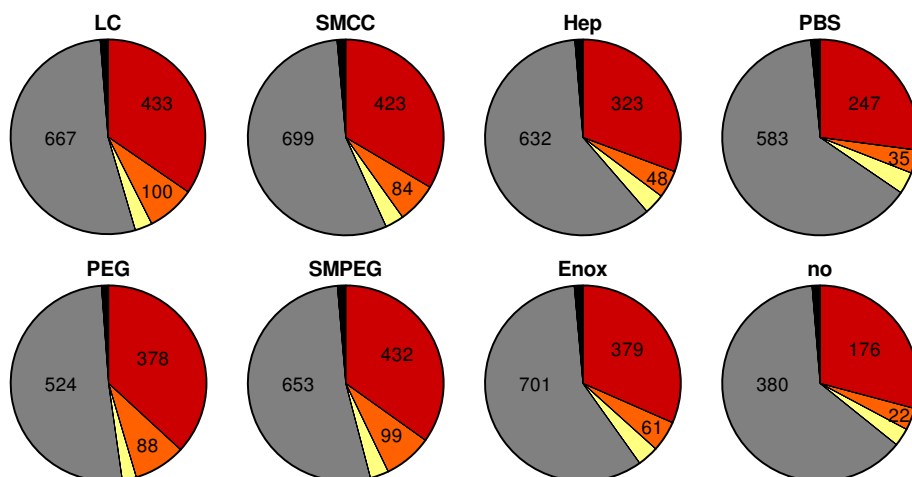


Figure 4.35.: Protein localization in the kidney dataset. Percentage and total protein numbers are depicted. Categories: Plasma membrane, extracellular and secreted (red), membrane associated (orange), internal side of plasma membrane (yellow), intracellular (grey), no annotation information available (black). Data without one hit wonders for the commercial reagents Sulfo-NHS-LC-biotin (LC) and NHS-PEG₁₂-biotin (PEG), the peptide-based reagents SMCC-Cys-(L-Asp)₃-biotin (SMCC) and SM(PEG)₆-Cys-(L-Asp)₃-biotin (SMPEG), the heparin-based reagents Biotin-Heparin-sNHS (Hep) and Biotin-Enoxaparin-sNHS (Enox) and the PBS- (PBS) or non-perfused (no) negative control is shown.

associated proteins constitute a 40-45% fraction with the commercial and peptide-based reagents. However, in terms of the numbers of identified proteins, no influence of the decreased reactivity of the peptide-based reagents compared to the commercial ones is observed. With the heparin-based reagents, a 35-37% enrichment of the desired proteome fraction is achieved. In the *in vivo* setting, no co-enrichment of nuclear proteins is detectable: The analysis of the subset Biotin-Heparin-sNHS vs. commercial Sulfo-NHS-LC-biotin revealed, that only 7% of the proteins identified with Biotin-Heparin-sNHS are exclusively identified within this group. An enrichment analysis of those proteins did not point towards any specifically enriched subcellular localization. Within the Biotin-Enoxaparin-sNHS vs. Sulfo-NHS-LC-biotin subset, the same is true for a 10% fraction of the proteins identified with Biotin-Enoxaparin-sNHS.

The closer analysis of the plasma membrane and extracellular matrix annotated fraction (see Table 4.5) revealed the comparability of the peptide-based reagents to the commercial reagents in terms of identified protein numbers, whereas the heparin-based reagents show reduced reactivity. Variability across different reagent groups is reduced if proteins identified with only one proteotypic peptide are excluded from analysis. Within the negative controls, the number of identified proteins annotated to the plasma membrane is reduced, but the fraction is still surprisingly high. Reasons are further discussed in chapter 4.4.4.2.

Subset	Protein number	LC & Sample	LC only	Sample only
LC & PEG	537	409 (76.2%)	116 (21.6%)	12 (2.2%)
LC & SMCC	577	433 (75.1%)	92 (15.9%)	52 (9.0%)
LC & SMPEG	586	439 (74.9%)	86 (14.7%)	61 (10.4%)
LC & Hep	561	328 (58.5%)	197 (35.1%)	36 (6.4%)
LC & Enox	594	376 (63.3%)	149 (25.1%)	69 (11.6%)
LC & PBS	550	245 (44.5%)	280 (50.9%)	25 (4.6%)
LC & no	531	176 (33.1%)	349 (65.7%)	6 (1.2%)
LC & PEG	436	375 (86.0%)	58 (13.3%)	3 (0.7%)
LC & SMCC	461	395 (85.7%)	38 (8.2%)	28 (6.1%)
LC & SMPEG	463	402 (86.8%)	31 (6.7%)	30 (6.5%)
LC & Hep	454	302 (66.5%)	131 (28.9%)	21 (4.6%)
LC & Enox	469	343 (73.1%)	90 (19.2%)	36 (7.7%)
LC & PBS	448	232 (51.8%)	201 (44.9%)	15 (3.3%)
LC & no	437	172 (39.4%)	261 (59.7%)	4 (0.9%)

Table 4.5.: Identification of proteins annotated to plasma membrane or extracellular matrix. Upper panel: all proteins, lower panel: without one hit wonders. Protein number and overlap of all sample groups against commercial Sulfo-NHS-LC-biotin (LC) in protein numbers and percentage. Commercial reagent NHS-PEG₁₂-biotin (PEG), peptide-based reagents SMCC-Cys-(L-Asp)₃-biotin (SMCC) and SM(PEG)₆-Cys-(L-Asp)₃-biotin (SM-PEG), heparin-based reagents Biotin-Heparin-sNHS (Hep) and Biotin-Enoxaparin-sNHS (Enox) and the PBS- (PBS) or non-perfused (no) negative controls.

Whereas in terms of identified protein numbers, the diversity among the different biotinylation groups and the controls are smaller than expected, differences occur at the level of the number of identified proteotypic peptides per protein (see Figure 4.36). High numbers of identified proteotypic peptides enable a more stable identification and quantification of the corresponding proteins. With commercial and peptide-based biotinylation reagents, the average number of identified proteotypic peptides per protein is significantly increased for the desired proteome fraction (plasma membrane, extracellular, membrane annotated), whereas the intracellular background proteins remain stable across the samples. Hereby, the highest average number of identified peptides per plasma membrane, extracellular or membrane annotated protein were obtained with the peptide-based reagents SM(PEG)₆-Cys-

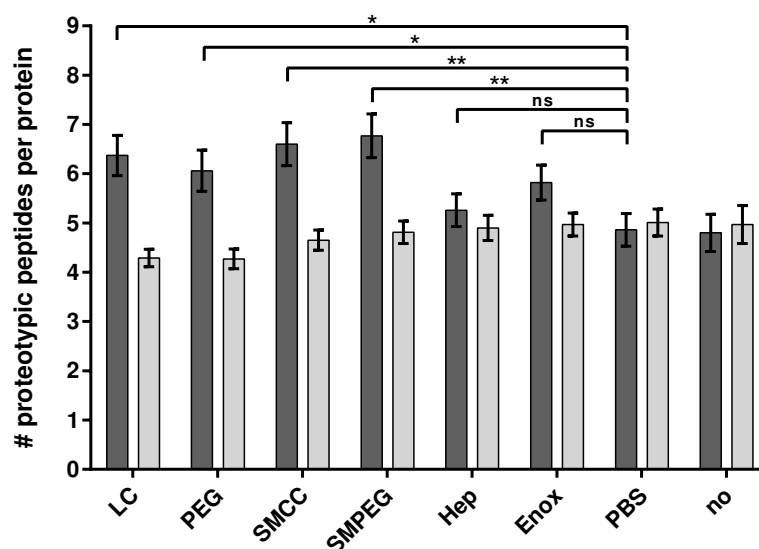


Figure 4.36.: Average number of proteotypic peptides per identified protein in the kidney dataset with SEM. Proteins annotated to the categories plasma membrane, extracellular and membrane in dark grey, intracellular proteins in light grey. Data includes all identified proteins for the commercial reagents Sulfo-NHS-LC-biotin (LC) and NHS-PEG₁₂-biotin (PEG), the peptide-based reagents SMCC-Cys-(L-Asp)₃-biotin (SMCC) and SM(PEG)₆-Cys-(L-Asp)₃-biotin (SMPEG), the heparin-based reagents Biotin-Heparin-sNHS (Hep) and Biotin-Enoxaparin-sNHS (Enox) and the PBS- (PBS) or non-perfused (no) negative controls. Unpaired t-test with significance levels ** ($P \leq 0.01$), * ($P \leq 0.05$), ns ($P > 0.05$).

(L-Asp)₃-biotin and SMCC-Cys-(L-Asp)₃-biotin.

Differences between the reagent and control groups, as well as between technical and biological replicates were further evaluated by relative quantification using the in-house developed software MS_QBAT.^[282] The method is further described in chapters 3.8.4 and 4.3.4.1. Data analysis included normalisation to internal standard peptides, annotation, feature extraction and complexity reduction: On average, 42% of the features per sample could be dismissed while keeping 99.5% of the identifications. Cross-alignments for propagation of annotations was performed as two-step procedure: first within pairs of technical replicates, secondly within all biological replicates belonging to the same reagent group. Variability between technical and biological replicates was further examined via pair-wise alignment by a genetic algorithm (GEAL alignment). For relative quantification, the eight samples belonging to one biotinylation reagent (four biological replicates with two technical replicates each) were combined to a super-sample by GEAL alignment, and quantified against the Sulfo-NHS-LC-biotin super-sample (see Suppl. Table B.7).

Pairwise quantification of technical and biological replicates to visualize the variability show a narrow distribution for the technical replicates, that broadens if biological replicates are compared. (see Figures 4.37 and 4.38). Even for the biological replicates 50% of all values are within a range of maximal 1.13-fold regulation.

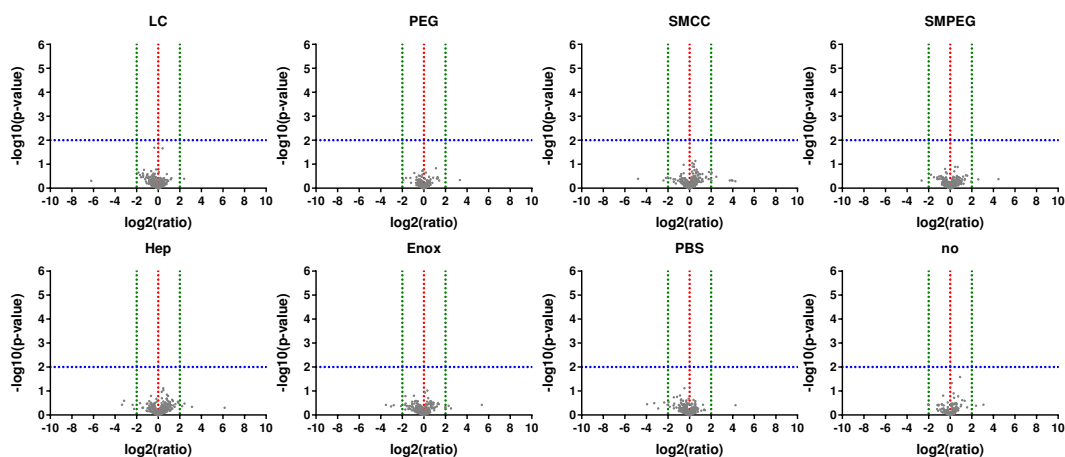


Figure 4.37.: Volcano plots of technical replicates (exemplary) for Sulfo-NHS-LC-biotin (LC), NHS-PEG₁₂-biotin (PEG), SMCC-Cys-(L-Asp)₃-biotin (SMCC), SM(PEG)₆-Cys-(L-Asp)₃-biotin (SMPEG), Biotin-Heparin-sNHS (Hep) and Biotin-Enoxaparin-sNHS (Enox) and the PBS- (PBS) or non-perfused (no) negative controls. Proteins quantified with at least two proteotypic peptides are depicted. No regulation ($x = 0$) and ± 4 -fold regulation ($x = \pm 2$) shown as vertical lines, a p-value of 0.01 as horizontal line ($y = 2$).

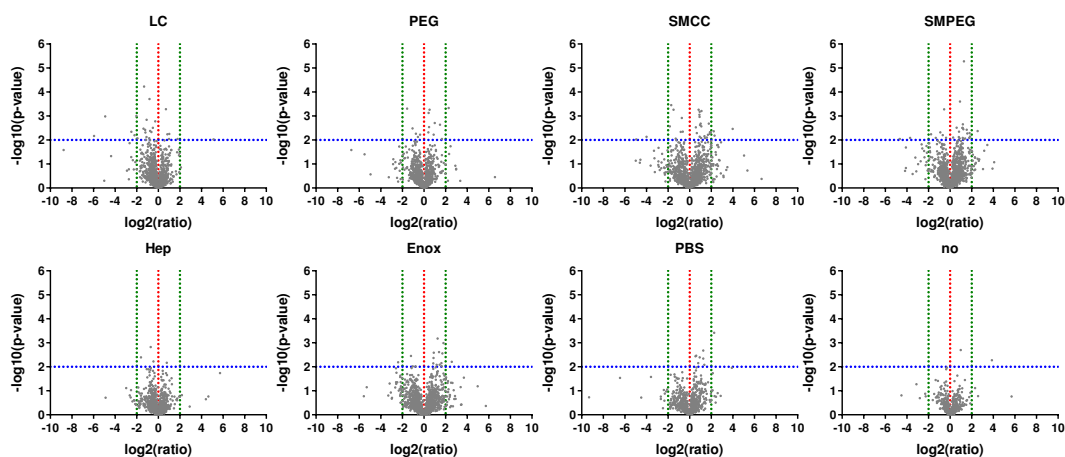


Figure 4.38.: Volcano plots of biological replicates (exemplary) for Sulfo-NHS-LC-biotin (LC), NHS-PEG₁₂-biotin (PEG), SMCC-Cys-(L-Asp)₃-biotin (SMCC), SM(PEG)₆-Cys-(L-Asp)₃-biotin (SMPEG), Biotin-Heparin-sNHS (Hep) and Biotin-Enoxaparin-sNHS (Enox) and the PBS- (PBS) or non-perfused (no) negative controls. Proteins quantified with at least two proteotypic peptides are depicted. No regulation ($x = 0$) and ± 4 -fold regulation ($x = \pm 2$) shown as vertical lines, a p-value of 0.01 as horizontal line ($y = 2$).

To compensate for biological and technical variation, all samples belonging to one reagent group were combined and relatively quantified to the Sulfo-NHS-LC-biotin super-sample. Volcano plots for those quantifications are depicted in Figure 4.39. Distributions are wide-spread and, especially in the negative controls, proteins are heavily down-regulated compared to Sulfo-NHS-LC-biotin.

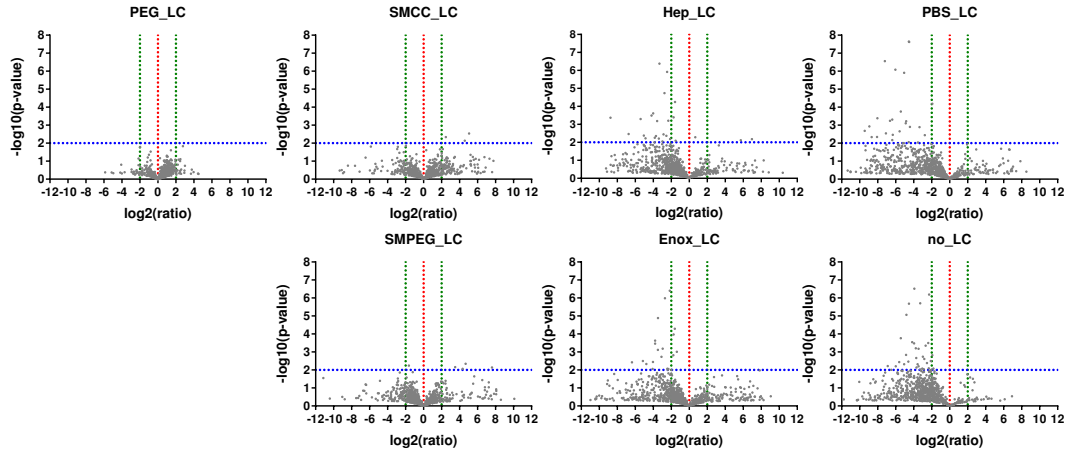


Figure 4.39.: Volcano plots of super-samples from all reagents (NHS-PEG₁₂-biotin (PEG), SMCC-Cys-(L-Asp)₃-biotin (SMCC), SM(PEG)₆-Cys-(L-Asp)₃-biotin (SMPEG), Biotin-Heparin-sNHS (Hep) and Biotin-Enoxaparin-sNHS (Enox) and the PBS- (PBS) or non-perfused (no) negative controls) against the Sulfo-NHS-LC-biotin (LC) super-sample. Proteins quantified with at least two proteotypic peptides are depicted. No regulation ($x = 0$) and ± 4 -fold regulation ($x = \pm 2$) shown as vertical lines, a p-value of 0.01 as horizontal line ($y = 2$).

To sum up the findings about the variability between different samples, the quantification of technical and biological replicates as well as of different reagent groups against Sulfo-NHS-LC-biotin are shown in a direct comparison (see Figure 4.40).

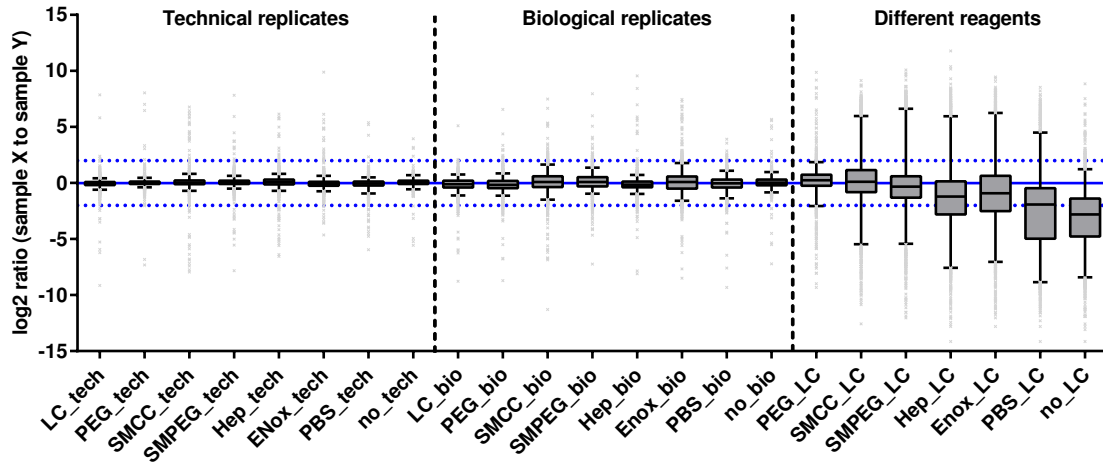


Figure 4.40.: Quantification of technical (tech) and biological replicates (bio) (exemplary) and all biotinylation reagents (NHS-PEG₁₂-biotin (PEG), SMCC-Cys-(L-Asp)₃-biotin (SMCC), SM(PEG)₆-Cys-(L-Asp)₃-biotin (SMPEG), Biotin-Heparin-sNHS (Hep), Biotin-Enoxaparin-sNHS (Enox) and PBS- (PBS) or non-perfused (no) negative controls) vs. Sulfo-NHS-LC-biotin (LC), all proteins. Box (median with 25th and 75th percentile) and whiskers (5th to 95th percentile) plot, outliers in grey. Blue straight line ($y = 0$) indicates no regulation, blue dotted lines ($y = \pm 2$) 4-fold regulation.

In total, 1851 proteins could be quantified, thereof 1242 proteins with at least two proteotypic peptides in one reagent group. Reactivity differences and down-regulation in the negative controls were further analysed by visualizing the median with the interquartile range (see Figure 4.41). Variation between both commer-

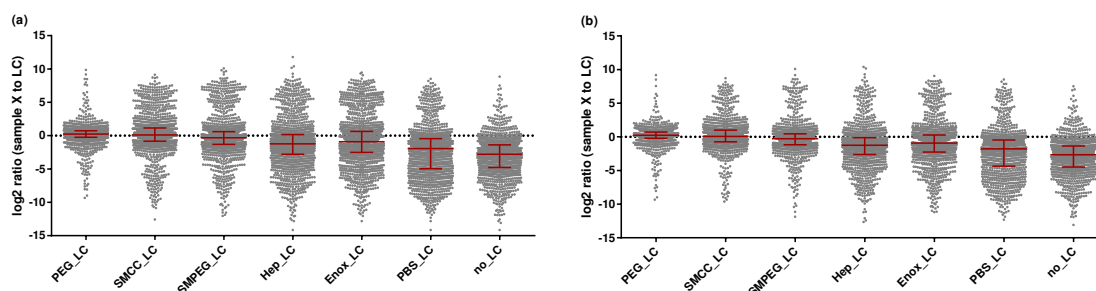


Figure 4.41.: Quantification of all biotinylation reagents (NHS-PEG₁₂-biotin (PEG), SMCC-Cys-(L-Asp)₃-biotin (SMCC), SM(PEG)₆-Cys-(L-Asp)₃-biotin (SMPEG), Biotin-Heparin-sNHS (Hep), Biotin-Enoxaparin-sNHS (Enox) and PBS- (PBS) or non-perfused (no) negative controls) vs. Sulfo-NHS-LC-biotin (LC), median with interquartile range. Blue dotted line ($y = 0$) indicates no regulation between samples. All proteins (a) or proteins quantified with at least two proteotypic peptides (b).

cial reagents NHS-PEG₁₂-biotin and Sulfo-NHS-LC-biotin is low. No reactivity-dependent shift is visible for the peptide-based reagents compared to Sulfo-NHS-LC-biotin. The less-reactive Biotin-Heparin-sNHS and Biotin-Enoxaparin-sNHS exhibit about two-fold down-regulation compared to commercial Sulfo-NHS-LC-biotin. If one hit wonders are excluded from analysis, outliers are removed and the distribution is narrowing. The non-perfused negative control is 7-fold down-regulated compared to Sulfo-NHS-LC-biotin, the PBS-perfused control, which is less complex due to blood removal, about 4-fold. Despite of the high background proteome identification in the negative controls, the proteins are clearly down-regulated compared to the biotinylated samples. In total, differences among the reagents' reactivity as well as the negative controls' down-regulation are less pronounced in the *in vivo* dataset than in the *in vitro* dataset (see chapter 4.3.4.1). This effect is first of all influenced by the higher background protein amount identified across all reagents in the *in vivo* dataset. Secondly, the amount of biotinylation reagent used for perfusion compensates better for reactivity differences than during *in vitro* biotinylation. Limiting factor during the *in vitro* biotinylation was the negative effect of higher reagent concentrations on the cell viability. For perfusion of mice, the reagent amount of commercial and peptide-based reagents was adapted to yield comparable results in terms of signal intensity in CLSM analysis (see chapter 4.4.2.2) and in terms of protein numbers in the proteomic dataset.

Cluster analysis of all proteins quantified against Sulfo-NHS-LC-biotin with at least two peptides revealed, that the similarity is both reactivity- and reagent-type dependent. In accordance with the previously determined decreased reactivity, the heparin-based reagents cluster with the negative controls. Hereby, the non-perfused

control is further apart than the PBS-perfused control, as the loss of blood components is a common characteristic for all perfused mice. Commercial and peptide-based reagents exhibit comparable reactivity in the *in vivo* dataset as shown before, and also cluster together. According to structural similarity of the biotinylation reagents, the peptide-based ones hereby form a sub-cluster.

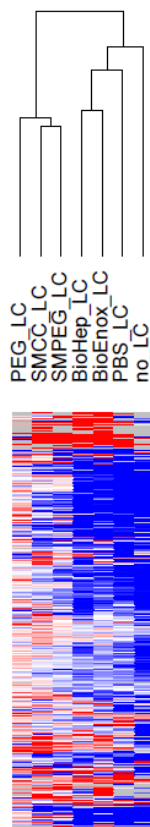


Figure 4.42.: Cluster analysis kidney dataset using GENE-E.^[285] Proteins quantified with at least two peptides. City block distance and average linkage, 4-fold up-regulation (red), 4-fold down-regulation (blue). All biotinylation reagents (NHS-PEG₁₂-biotin (PEG), SMCC-Cys-(L-Asp)₃-biotin (SMCC), SM(PEG)₆-Cys-(L-Asp)₃-biotin (SMPEG), Biotin-Heparin-sNHS (BioHep), Biotin-Enoxaparin-sNHS (BioEnox)) and PBS- (PBS) or non-perfused (no) negative controls vs. Sulfo-NHS-LC-biotin (LC).

Finally, the desired subset with plasma membrane or extracellular matrix annotation was analysed (see Figure 4.43). Commercial and peptide-based reagents are comparable in terms of the quantified signal intensity of the desired proteome fraction. For the heparin-based reagents, the plasma membrane/extracellular proteome fraction is down-regulated. One reason is the introduction of a background intensity value in a presence vs. absence situation, leading to an artificially high ratio, rather than removal of this protein as non-quantifiable. Secondly, also for the stably identified proteins, the signal intensity is lower with the heparin-based

reagents due to their decreased reactivity. Negative control signal intensity is significantly decreased in the plasma membrane/extracellular fraction. Proteins are more than ten-fold down-regulated within the negative controls, which is a ratio pointing towards a presence vs. absence situation in the majority of the proteins.

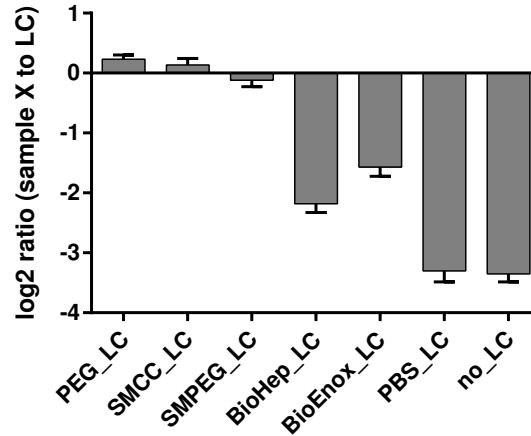


Figure 4.43.: Analysis of all proteins with plasma membrane or extracellular annotation quantified with at least two peptides. Quantification of all samples (NHS-PEG₁₂-biotin (PEG), SMCC-Cys-(L-Asp)₃-biotin (SMCC), SM(PEG)₆-Cys-(L-Asp)₃-biotin (SMPEG), Biotin-Heparin-sNHS (Hep), Biotin-Enoxaparin-sNHS (Enox) and PBS- (PBS) or non-perfused (no) negative controls vs. Sulfo-NHS-LC-biotin (LC). Median with SEM.

A detailed analysis of protein regulation within the subsets (see Figure 4.44) shows the high similarity of both commercial reagents Sulfo-NHS-LC-biotin and NHS-PEG₁₂-biotin: 71% of all plasma membrane and extracellular matrix proteins are within a range of ± 2 -fold regulation. Peptide-based reagents in comparison to Sulfo-NHS-LC-biotin exhibit a larger variation: About 40% of the quantified

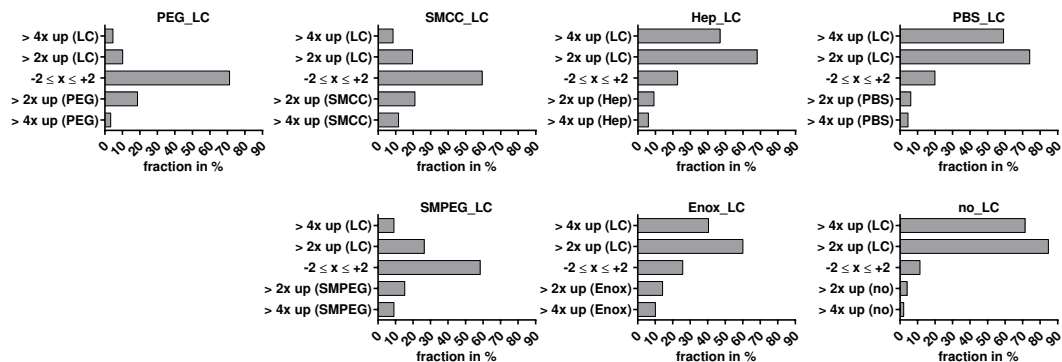


Figure 4.44.: Detailed regulation analysis of all proteins with plasma membrane or extracellular annotation quantified with at least two peptides. Quantification of all samples (NHS-PEG₁₂-biotin (PEG), SMCC-Cys-(L-Asp)₃-biotin (SMCC), SM(PEG)₆-Cys-(L-Asp)₃-biotin (SMPEG), Biotin-Heparin-sNHS (Hep), Biotin-Enoxaparin-sNHS (Enox) and PBS- (PBS) or non-perfused (no) negative controls vs. Sulfo-NHS-LC-biotin (LC).

proteins are more than 2-fold up- or down-regulated, about 20% more than 4-fold. Within the heparin-based reagents Biotin-Heparin-sNHS and Biotin-Enoxaparin-sNHS the reactivity shift is visible: 68% and 60% of the quantified proteins are more than 2-fold up-regulated within the Sulfo-NHS-LC-biotin group, respectively. The desired plasma membrane and extracellular matrix proteome fraction is severely down-regulated within the negative controls.

The study of *in vivo* biotinylated kidney samples demonstrated comparability in enrichment of the desired proteome fraction of commercial and peptide-based reagents, both in terms of identified total protein numbers and in signal intensities of the proteotypic peptides. Although a reduced reactivity of the peptide-based reagents was shown previously, the quality of the enrichment of the desired proteome fraction was not influenced in the *in vivo* setting. The enrichment of nuclear proteins with the heparin-based reagents had no effect during *in vivo* perfusion of mice. Nevertheless, the reactivity was decreased compared to other biotinylation reagents. The high number of background proteins identified within the negative controls will be further elucidated in the following chapter. The interplay between selectivity and reactivity is the challenge for the work with novel biotinylation reagents.

4.4.4.2. Background Proteome in Negative Controls

As shown in chapter 4.4.4.1, most proteins within the PBS- or unperfused negative controls, are severely down-regulated in terms of signal intensity in comparison to the Sulfo-NHS-LC-biotin sample. Nevertheless, 1008 (912) and 633 (597) proteins have been identified in the PBS- and in the non-perfused control, respectively, numbers, that are surprisingly high compared to the *in vitro* negative control (see chapter 4.3.4.1).

To elucidate possible reasons for background protein binding, different sample preparation strategies were examined: (i) Washing with high-salt buffer (2 M NaCl) to disrupt non-covalent protein-protein interactions, (ii) pre-blocking of the streptavidin-sepharose with BSA to exclude non-specific protein binding, (iii) Ultracentrifugation of the tissue homogenates to exclude the transfer of protein-containing particles to the streptavidin-sepharose, that would be co-purified on the filter and therefore be co-digested with the proteins on the beads and (iv) benzonase nuclease treatment after protein precipitation to degrade nucleic acid and to release nuclear proteins, that could be co-purified with the streptavidin-sepharose on the filter. All sample preparation strategies included reduction of disulfide bridges with TCEP and alkylation with IAA, as it is reported to improve tryptic digest, sequence coverage and the detection of cysteine-containing peptides in standard proteomic workflows.^{[348],[349]} 2.5 mg of total protein were captured on 100 µl streptavidin-sepharose, samples were processed as described above. Proteins derived from 2 mg total protein amount were examined per LC/MS run. Database searches and extraction of proteotypic peptides were performed as described in chapter 4.4.4.1. To account for mass spectrometric performance differences, annotations were propagated across samples biotinylated with the same reagent with MS_QBAT (see Suppl. Table B.8).

No significant change of the fraction of identified intracellular proteins was obtained with any strategy (see Table 4.6). The number of identified proteins was reduced in precipitated, benzonase-treated samples, which however was first of all an effect of non-complete protein resolubilisation.

Sample	Preparation	Protein number	% intracellular
PBS	reduction/alkylation	935	62.7%
PBS	reduction/alkylation/high-salt	922	62.6%
PBS	preblocking/reduction/alkylation	918	62.8%
PBS	ultracentrifugation/reduction/alkylation	855	64.0%
PBS	precipitation/benzonase/reduction/alkylation	670	63.1%
no	reduction/alkylation	806	63.9%
no	reduction/alkylation/high-salt	785	63.8%
no	preblocking/reduction/alkylation	815	63.5%
no	ultracentrifugation/reduction/alkylation	820	63.0%
no	precipitation/benzonase/reduction/alkylation	745	64.1%
NHS-PEG ₁₂ -biotin	reduction/alkylation	1256	47.9%
NHS-PEG ₁₂ -biotin	reduction/alkylation/high-salt	1248	47.5%
NHS-PEG ₁₂ -biotin	preblocking/reduction/alkylation	1265	47.8%
NHS-PEG ₁₂ -biotin	ultracentrifugation/reduction/alkylation	1246	47.4%
NHS-PEG ₁₂ -biotin	precipitation/benzonase/reduction/alkylation	1151	46.2%

Table 4.6.: Sample preparation tests on kidney negative control samples. Perfusion with PBS (PBS) or non-perfused (no), perfusion with NHS-PEG₁₂-biotin as control. Number of identified proteins (without one hit wonders) and percentage of the intracellular fraction are shown. Sample preparation strategies are further described in the text.

Delipidation with organic solvents (1-butanol/diisopropylether) is a technique reported for efficient release of plasma membrane proteins from the surrounding lipids.^[347] The delipidation step was included for both kidney and liver samples, as it heavily facilitated sample preparation of tissue samples.

The identification of large background protein numbers may originate from a combination of different sources. Compared to the *in vitro* biotinylated samples, not only sample complexity, but also protein amount is increased: The total protein amount of the *in vivo* biotinylated samples loaded on the same amount of streptavidin-sepharose is 2.5-fold (kidney dataset) or 5-fold (negative control tests and liver dataset) higher compared to the capturing of the *in vitro* biotinylated samples. Unspecific binding to the streptavidin-sepharose resin plays a minor role, as only a small number of proteins was identified within the *in vitro* negative control and pre-blocking of the beads did not significantly influence the result. Still, background proteome binding might be higher due to the extensive and highly-concentrated total protein load in the *in vivo* situation. Furthermore, some strong non-covalent protein-protein interactions might not be disrupted by extensive washing. As elucidated in chapter 4.3.4.1, protein localization annotation in the database might be incomplete or wrong. The most influential explanation originates from an experimental observation: Tissue samples contain a vast amount of lipids. Even

after tissue homogenisation, sonication, heating in lysis buffer and centrifugation at $20\,000\times g$, the lysate cannot be completely cleared from lipids. Without the delipidation step, the separation of the streptavidin beads and the supernatant is hindered as the filter is often blocked by a lipid layer. During delipidation, an extensive lipid layer forms at the interface between the organic and the aqueous solvent. Although a vast lipid amount is removed by the procedure, it is experimentally barely possible to separate the streptavidin resin, the lipid layer and the solvents perfectly. Since the desired protein fraction is also not in solution, but bound to the streptavidin resin, some lipid structures are always subjected to tryptic digest with the beads. Co-digestion of a vast amount of non-biotinylated proteins, that are non-covalently bound to the lipid structures, is the result.

To further assess this hypothesis, the nature of the co-enriched background proteins was examined. An enrichment analysis of the top-300 proteins (identification with at least five proteotypic peptides) identified in the PBS-perfused negative control of the kidney dataset shows that most of those proteins belong to the severely enriched categories *membrane bounded organelle* and *mitochondrion* and are heavily involved in *catabolic processes* (see Figure 4.45). A pure protein abundance effect can be excluded, as no single protein of the 46 high-abundant intracellular ribosome components or proteasome subunits, that have been identified across the kidney dataset, are within the top-300 proteins of the PBS control. Therefore, the co-digestion of top-300 proteins cannot only originate from the high abundance of metabolic mitochondrial enzymes. Reason for the co-digestion of mitochondrial membrane bound proteins is more likely rooted in their lipid-binding domains, leading to their co-purification with lipid impurities. Furthermore, the identified plasma membrane annotated fraction within the negative controls was relatively high (18%), also pointing towards non-specific co-enrichment by lipid-binding.

Despite the relatively high identified protein numbers, the background proteome is significantly down-regulated in the kidney dataset as shown by relative quantification based on MS1 signal intensities, and does not hinder successful enrichment of the desired vascular accessible surface proteome fraction with the biotinylation reagents. Nevertheless, the existence of a PBS-perfused negative control in biomarker discovery studies might help to evaluate (by numbers of identified proteotypic peptides or relatively quantified signal intensities) if potential biomarkers have been identified specifically by enrichment of vascular accessible proteins or by non-specific co-purification.

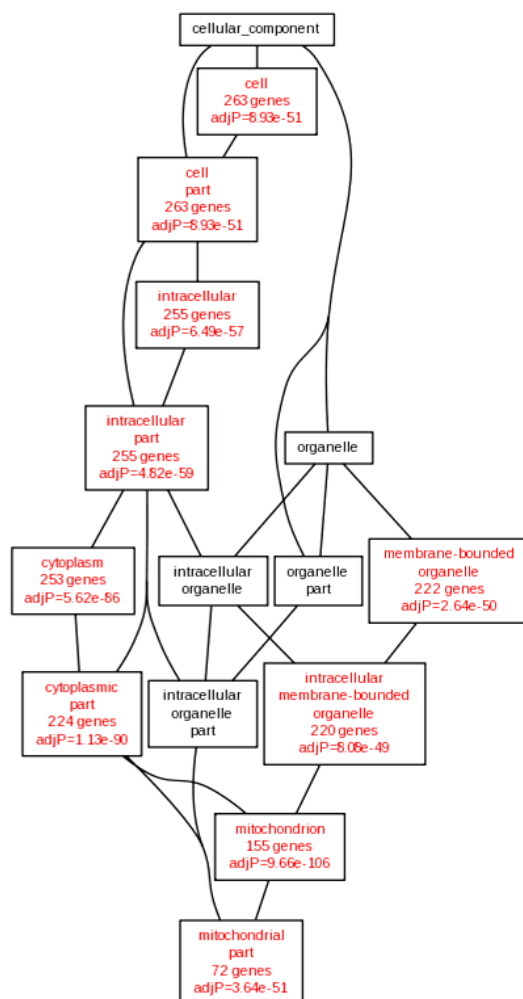


Figure 4.45.: Enrichment analysis of the cellular component of the top-300 proteins identified in PBS-perfused mouse kidney. Top-300 selection based on number of identified proteotypic peptides. Visualisation as directed acyclic graph using Webgestalt.^{[283],[284]} Enriched GO categories are depicted in red, name of the GO category, number of included genes and adjusted p-value are shown.

4.4.4.3. SRM-Analysis of High- and Low-Abundant Proteins

Protein candidates with intracellular, plasma membrane or extracellular localization annotation were chosen based on the protein lists from the discovery experiment. Proteotypic peptides were restricted to a mass range of 600-2000 Da, methionine and cysteine containing peptides were excluded if possible. Peptide standards corresponding to the natural counterparts, but labelled with heavy isotopic lysine or arginine were spiked into the peptide samples to enable relative quantification. For the final experiment, the three most intense transitions (charge state +2 or +3) per peptide were selected. SRM method optimisation and the quantification experiment were performed on a QTRAP[®] 6500 system (AB SCIEX).

Four biological replicates were measured per biotinylation reagent group (commercial reagents Sulfo-NHS-LC-biotin and NHS-PEG₁₂-biotin, peptide-based reagents SMCC-Cys-(L-Asp)₃-biotin and SM(PEG)₆-Cys-(L-Asp)₃-biotin, heparin-based reagents Biotin-Heparin-sNHS and Biotin-Enoxaparin-sNHS, negative control PBS- or non-perfused). Samples were prepared by capturing biotinylated proteins out of equal total protein amounts under reductive conditions, followed by alkylation, delipidation, tryptic digest and desalting, in accordance to the discovery experiment. To assess if the proteins are comparably targetable by SRM-based quantification with non-enriched full proteome samples, PBS- and non-perfused tissue homogenates were reduced, alkylated and digested after protein precipitation. Hereby, the protein amount subjected to analysis was adapted to exploit the full capacity of the LC/MS system.

To correct for mass spectrometric performance differences and ionization or spray differences between runs, the quantification was based on the ratio of the background-reduced peak areas from the light transitions and the background-reduced peak areas from the heavy transitions. Areas from three transitions per proteotypic peptide were summed for analysis. Peptides with unfavorable elution profile or interfering noise were excluded from analysis. The stability across the LC/MS runs was confirmed by a steady signal with low variation of the spiked heavy peptides (see Suppl. Figure A.10a,b). To obtain a reliable quantification, peptides with an average peak area less than four-fold up-regulated compared to the background signal were excluded from analysis. Enriched samples and full proteome samples were evaluated separately. Peak areas were normalised to the maximal peak area achieved among the enriched surface samples, mean and standard deviation were calculated across four biological replicates (see Figure 4.46).

The precipitated full proteome samples are not directly comparable to the enriched samples generated by capturing of the biotinylated proteins: The protein amount subjected to analysis is not equal, as the amount of vascular labelled biotinylated protein in the enriched samples is unknown and normalized to the same total protein amount subjected to capturing. Moreover, different sample preparation strategies come along with different percentages of sample loss. Yet, the full proteome analysis provides the information, if the targeted proteins are also stably quantifiable without prior enrichment.

Information about protein abundance in mouse kidney tissue was extracted from PaxDB, a protein abundance database based on publicly available experimental datasets.^{[286],[287]} The integrated and re-processed kidney protein abundance information from several publications was used.^{[280],[371],[372],[373]}

To investigate, if the stable identification of highly abundant intracellular proteins is decreased within the biotinylated samples compared to the full proteome samples, four protein candidates were selected (see Figure 4.46a-d): Calreticulin, a calcium-binding chaperone located in the endoplasmic reticulum lumen, as well as 60 kDa heat shock protein involved in correct folding of imported proteins in the mitochondrion matrix and the multifunctional nucleolar phosphoprotein Nucleolin, reported to play roles in chromatin structure regulation and the biogenesis of rRNA, all be-

long to the top 5% kidney proteins in terms of protein abundance.^{[374],[375],[376],[377]} The fourth selected protein, the 40S ribosomal protein S8, a component of the 40S ribosomal subunit, is reported to be amongst the top 25% abundant proteins.^[378] All intracellular proteins were very stably identified among the full proteome samples, whereas the signals across the biotinylated samples are low - despite of the high protein abundance.

Intracellular carboxylases are co-enriched with streptavidin-sepharose as biotin is a covalently bound cofactor required in the enzymatic reaction. Three carboxylases, all being at least amongst the top 25% abundant kidney proteins, have been selected to demonstrate the stable identification across all biotinylation reagents (see Figure 4.46e-g): Acetyl-CoA carboxylase, involved in long-chain fatty acid synthesis, Methylcrotonyl-CoA carboxylase, crucial for leucine and isovalerate catabolism, and Propionyl-CoA carboxylase, which is essential for the catabolism of odd-numbered fatty acid chains and of methyl-branched amino acids.^[369] The latter two are amongst the top 5% abundant kidney proteins, Acetyl-CoA carboxylase amongst the top 25%. Quantification works stably with the full proteome samples. As expected from the stable identification within the mass spectrometric discovery dataset, the biotin-dependent enzymes are severely enriched with the capturing approach.

Several cell surface or extracellular matrix proteins with different protein abundances reaching from the top 25% to the bottom 5% of the kidney proteome have been selected (see Figure 4.46h-o). The cell adhesion protein Cadherin-1, involved in the formation of adherens junctions of epithelial cells, and one of the major extracellular matrix components, the Collagen alpha-1(VI) chain, reported to be deregulated in metastatic breast cancer, belong to the top 25% abundant kidney proteins.^{[379],[380],[381]} Quantification of both proteins out of full lysates is possible, peptide-based reagents and Sulfo-NHS-LC-biotin exhibit comparable enrichment.

Next, three proteins with medium abundance levels are shown: the cell surface proteoglycan Glypican-4 involved in the development of kidney tubules, the immunoglobulin-like cell-cell adhesion molecule Nectin-3 as well as Integrin alpha-V, an Integrin receptor subunit, which is interacting with extracellular matrix components.^{[382],[383],[384]} None of those proteins could be quantified within the full proteome samples.

Finally, the detection of low-abundant proteins was examined: The cell adhesion molecule Cadherin-2 is mostly expressed in the nervous system and belongs to the bottom 25% proteins in terms of kidney protein abundance.^[379] Integrin beta-5 is among the bottom 5%. Receptor-type tyrosin phosphatase U regulating β -Catenin function both in adhesion and signalling, was not identified in the kidney datasets integrated into PaxDB, in terms of whole mouse body abundance it belongs to the bottom 25%.^[385] SRM-based quantification was possible with the biotinylated samples, but not with the full proteome samples. Proteins are down-regulated in the heparin-based reagent samples in accordance to their low reactivity. Both commercial and peptide-based reagents provide a stable technology platform for the quantification of low-abundant surface or extracellular matrix annotated biomarker candidates.

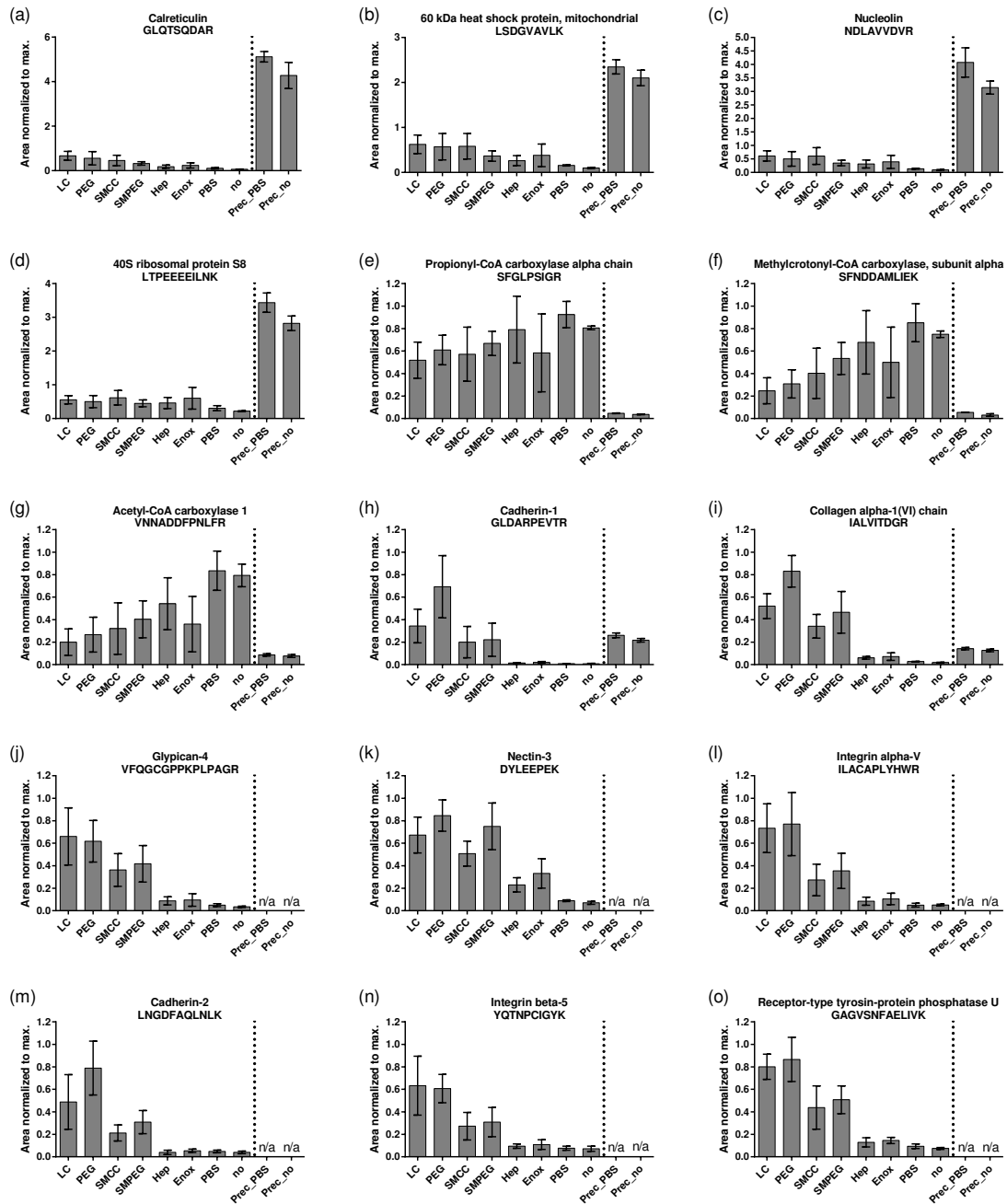


Figure 4.46.: SRM analysis of kidney samples, mean with standard deviation of four biological replicates. Vascular accessible proteome enrichment with commercial reagents Sulfo-NHS-LC-biotin (LC) and NHS-PEG₁₂-biotin (PEG), peptide-based reagents SMCC-Cys-(L-Asp)₃-biotin (SMCC) and SM(PEG)₆-Cys-(L-Asp)₃-biotin (SMPEG), heparin-based reagents Biotin-Heparin-sNHS (Hep) and Biotin-Enoxaparin-sNHS (Enox), negative control PBS- (PBS) or non-perfused (no). Full proteome samples of precipitated proteins from PBS- (Prec_PBS) or non-perfused (Prec_no) mice. Proteins with intracellular annotation (a-d), carboxylases (e-g), cell membrane or extracellular matrix annotation (h-o).

4.4.5. Mass Spectrometric Analysis of Biotinylated Liver Tissue

4.4.5.1. Discovery Experiment Liver

For comparative proteomic analysis, capturing of biotinylated proteins was performed with equal amounts of 2.55 mg total liver protein on 100 μ L streptavidin-sepharose. Samples were processed by reduction, alkylation, delipidation and tryptic digest. As a lower biotinylation degree of liver tissue compared to kidney tissue was detected by CLSM and ELISA, peptides derived from 1.25 mg total protein amount (kidney: 750 μ g) were examined per LC/MS run. Three biological replicates per reagent group (commercial reagents: Sulfo-NHS-LC-biotin, NHS-PEG₁₂-biotin, peptide-based reagents: SMCC-Cys-(L-Asp)₃-biotin, SM(PEG)₆-Cys-(L-Asp)₃-biotin, heparin-based reagents: Biotin-Heparin-sNHS, Biotin-Enoxaparin-sNHS, negative controls: PBS- or non-perfused) were analysed.

Consistent chromatographic separation across the samples was visualised by 2D peptide mapping (see Suppl. Figure A.9). The analysis of MS data based on proteotypic peptide summaries was performed as already described in chapter 4.4.4.1.

In total, 1531 proteins were identified, thereof 1120 proteins with more than one proteotypic peptide in at least one reagent group (without one hit wonders) (see Suppl. Table B.9). 1102 (971) proteins were found with commercial Sulfo-NHS-LC-biotin, 867 (804) with NHS-PEG₁₂-biotin. With the novel peptide-based reagent SMCC-Cys-(L-Asp)₃-biotin 1281 (1034) proteins were obtained, 1097 (966) with SM(PEG)₆-Cys-(L-Asp)₃-biotin. 945 (853) proteins were identified with Biotin-Heparin-sNHS, 1000 (882) with Biotin-Enoxaparin-sNHS. 45% (51%) of all identified proteins within the dataset were found with every biotinylation reagent. The overlap fraction between subsets of two sample groups is large and independent of the nature of the biotinylation reagent, e.g. 76% (84%) between Sulfo-NHS-LC-biotin and SMCC-Cys-(L-Asp)₃-biotin, 74% (84%) between Sulfo-NHS-LC-biotin and Biotin-Heparin-sNHS and 71% (79%) between both peptide-based reagents. Within the PBS- or unperfused negative controls, 1003 (879) and 807 (740) proteins have been identified respectively. The identified background proteome will be discussed later in this chapter.

Differences in protein localization across the various biotinylation reagents was analysed via extraction of subcellular localization information from the UniProt website. The localization information was curated by screening for the categories plasma membrane, extracellular and secreted, internal side of the plasma membrane, membrane associated, intracellular and no information available (see Figure 4.47). Compared to the kidney dataset, the surface/extracellular matrix fraction is reduced. Proteins annotated to plasma membrane, the extracellular matrix and membrane constitute a 27-34% fraction with the commercial and peptide-based reagents, heparin-based reagents and the negative controls exhibit both similar proteins numbers and fractions.

Also at the level of the number of identified proteotypic peptides per protein, differences between samples and negative controls are small (see Figure 4.48). The

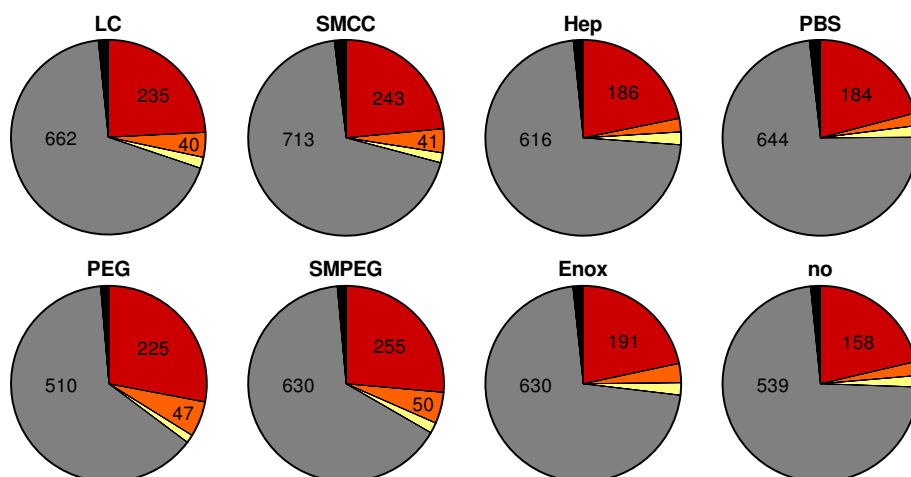


Figure 4.47.: Protein localization in the liver dataset. Percentage and total protein numbers are depicted. Categories: Plasma membrane, extracellular and secreted (red), membrane associated (orange), internal side of plasma membrane (yellow), intracellular (grey), no annotation information available (black). Data without one hit wonders for the commercial reagents Sulfo-NHS-LC-biotin (LC) and NHS-PEG₁₂-biotin (PEG), the peptide-based reagents SMCC-Cys-(L-Asp)₃-biotin (SMCC) and SM(PEG)₆-Cys-(L-Asp)₃-biotin (SMPEG), the heparin-based reagents Biotin-Heparin-sNHS (Hep) and Biotin-Enoxaparin-sNHS (Enox) and the PBS- (PBS) or non-perfused (no) negative control is shown.

number of proteotypic peptides identified for the plasma membrane/extracellular fraction is slightly, but non-significantly reduced within the negative controls (unpaired t-test, $0.10 \leq P \leq 0.20$ for the comparisons SMPEG vs. PBS and SMCC vs. PBS).

Relative quantification was performed using the in-house developed software MS_QBAT.^[282] The method is further described in chapters 3.8.4 and 4.3.4.1. Data analysis included normalisation against internal standard peptides, annotation, feature extraction and complexity reduction: On average, 47% of the features per sample could be dismissed while keeping 99.5% of the identifications. Annotations were propagated after cross-alignment of all pairs of biological replicates belonging to the same reagent group. For relative quantification, three biological replicates belonging to one biotinylation reagent were combined to a super-sample by GEAL alignment, and quantified against the Sulfo-NHS-LC-biotin super-sample (see Suppl. Table B.10).

Variability between biological replicates was further examined via pair-wise GEAL alignment. As expected, the variability between biological replicates is lower compared to the variability between different reagents (see Figures 4.49, 4.50).

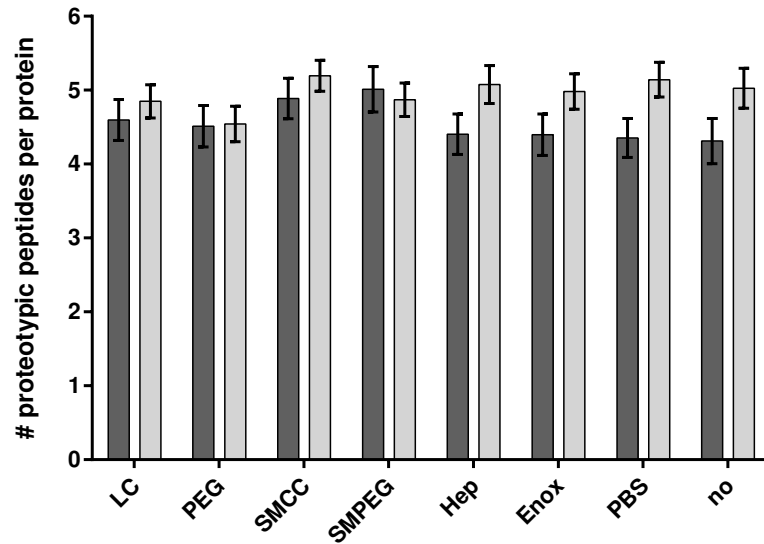


Figure 4.48.: Average number of proteotypic peptides per identified protein in the liver dataset with SEM. Proteins annotated to the categories plasma membrane, extracellular and membrane in dark grey, intracellular proteins in light grey. Data includes all identified proteins for the commercial reagents Sulfo-NHS-LC-biotin (LC) and NHS-PEG₁₂-biotin (PEG), the peptide-based reagents SMCC-Cys-(L-Asp)₃-biotin (SMCC) and SM(PEG)₆-Cys-(L-Asp)₃-biotin (SMPEG), the heparin-based reagents Biotin-Heparin-sNHS (Hep) and Biotin-Enoxaparin-sNHS (Enox) and the PBS- (PBS) or non-perfused (no) negative controls.

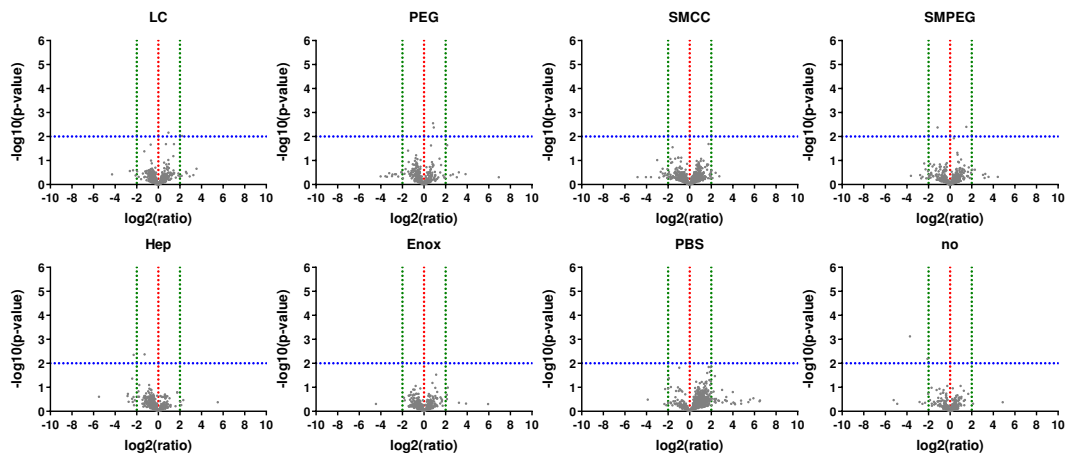


Figure 4.49.: Volcano plots of biological replicates (exemplary) for Sulfo-NHS-LC-biotin (LC), NHS-PEG₁₂-biotin (PEG), SMCC-Cys-(L-Asp)₃-biotin (SMCC), SM(PEG)₆-Cys-(L-Asp)₃-biotin (SMPEG), Biotin-Heparin-sNHS (Hep) and Biotin-Enoxaparin-sNHS (Enox) and the PBS- (PBS) or non-perfused (no) negative controls. Proteins quantified with at least two proteotypic peptides are depicted. No regulation ($x = 0$) and ± 4 -fold regulation ($x = \pm 2$) shown as vertical lines, a p-value of 0.01 as horizontal line ($y = 2$).

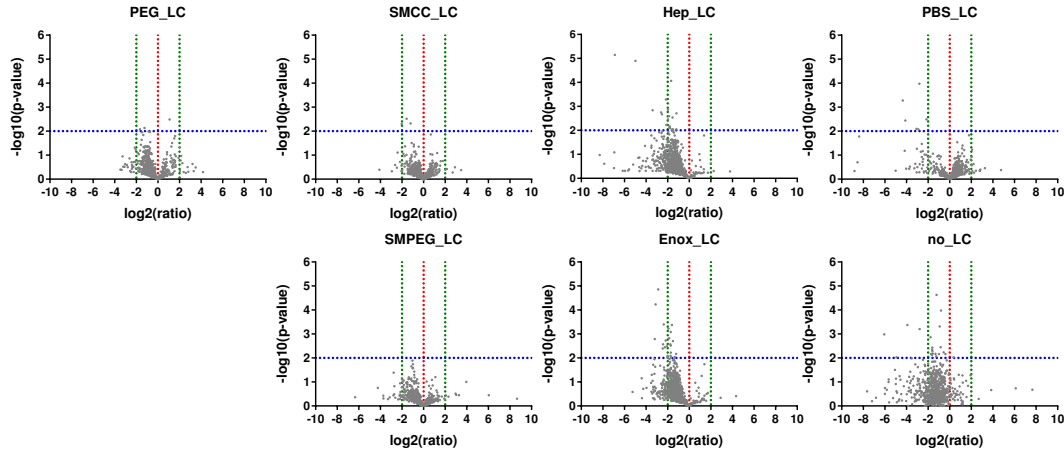


Figure 4.50.: Volcano plots of super-samples from all reagents (NHS-PEG₁₂-biotin (PEG), SMCC-Cys-(L-Asp)₃-biotin (SMCC), SM(PEG)₆-Cys-(L-Asp)₃-biotin (SMPEG), Biotin-Heparin-sNHS (Hep) and Biotin-Enoxaparin-sNHS (Enox) and the PBS- (PBS) or non-perfused (no) negative controls) against the Sulfo-NHS-LC-biotin (LC) super-sample. Proteins quantified with at least two proteotypic peptides are depicted. No regulation ($x = 0$) and ± 4 -fold regulation ($x = \pm 2$) shown as vertical lines, a p-value of 0.01 as horizontal line ($y = 2$).

In total, 1458 proteins could be quantified, thereof 1026 proteins with at least two proteotypic peptides in one reagent group. General differences in the signal intensities were analysed by visualizing the median with the interquartile range (see Figure 4.51). The distribution is narrowing, when only proteins quantified with at least two proteotypic peptides are included. Compared to the kidney dataset, the regulation between the different reagent groups is small. A maximum of 2.5-fold down-regulation is visible in the heparin-based reagent group. Proteins quantified within the non-perfused negative control are slightly down-regulated compared to the PBS-perfused control as also seen in the kidney dataset. Reason is the reduced

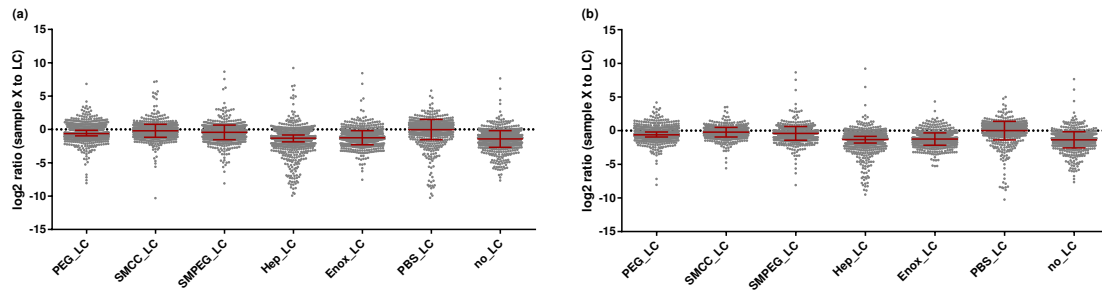


Figure 4.51.: Quantification of all biotinylation reagents (NHS-PEG₁₂-biotin (PEG), SMCC-Cys-(L-Asp)₃-biotin (SMCC), SM(PEG)₆-Cys-(L-Asp)₃-biotin (SMPEG), Biotin-Heparin-sNHS (Hep), Biotin-Enoxaparin-sNHS (Enox) and negative control (PBS)) vs. Sulfo-NHS-LC-biotin (LC), median with interquartile range. Blue dotted line ($y = 0$) indicates no regulation between samples. All proteins (a) or without one hit wonders (b).

sample complexity due to the removal of blood components during perfusion. Nevertheless, the negative controls, especially the PBS-perfused one, exhibit only small differences to Sulfo-NHS-LC-biotin. Reasons will be discussed later in this chapter.

To reveal similarities among the reagent groups, cluster analysis was performed based on all proteins quantified against Sulfo-NHS-LC-biotin with at least two peptides (see Figure 4.52). Similar to the kidney dataset, structural similarity of the biotinylation reagents determines proximity in the sub-clusters. Nevertheless, differences to the negative controls are smaller than in the kidney dataset.

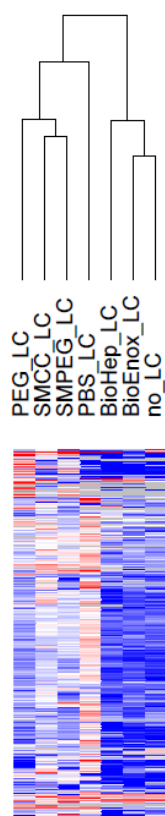


Figure 4.52.: Cluster analysis liver dataset using GENE-E.^[285] Proteins quantified with at least two peptides. City block distance and average linkage, 4-fold up-regulation (red), 4-fold down-regulation (blue). All biotinylation reagents (NHS-PEG₁₂-biotin (PEG), SMCC-Cys-(L-Asp)₃-biotin (SMCC), SM(PEG)₆-Cys-(L-Asp)₃-biotin (SMPEG), Biotin-Heparin-sNHS (BioHep), Biotin-Enoxaparin-sNHS (BioEnox)) and PBS- (PBS) or non-perfused (no) negative controls vs. Sulfo-NHS-LC-biotin (LC).

Liver samples exhibit a drastically reduced surface/extracellular matrix fraction compared to the kidney samples. Across all analysis performed on the dataset, the high similarity across all samples and the negative controls both in terms of protein identification and in relative quantification is striking. In contrast to the kidney samples, the total protein amount loaded on the streptavidin-sepharose was higher, as

liver tissue biotinylation is lower (shown by ELISA and CLSM analysis). One factor for the identification of a large background proteome might be non-specific protein binding to the streptavidin-sepharose. In fact, liver homogenates were even higher concentrated than kidney homogenates, due to the reduced biotinylated proteome fraction. Furthermore, the extremely high lipid content within the liver samples was obvious during sample preparation. As elucidated in chapter 4.4.4.2, co-digestion of lipid-bound proteins with the biotin-labelled and streptavidin-bound proteins is likely to occur in tissue samples. To apply the hypothesis to the liver dataset, an enrichment analysis of 605 proteins identified with all reagents and both negative control groups was performed (see Figure 4.53). *Membrane-bounded organelle* is heavily enriched as subcellular localization, pointing towards a co-enrichment of high-abundant proteins with lipid-binding domains.

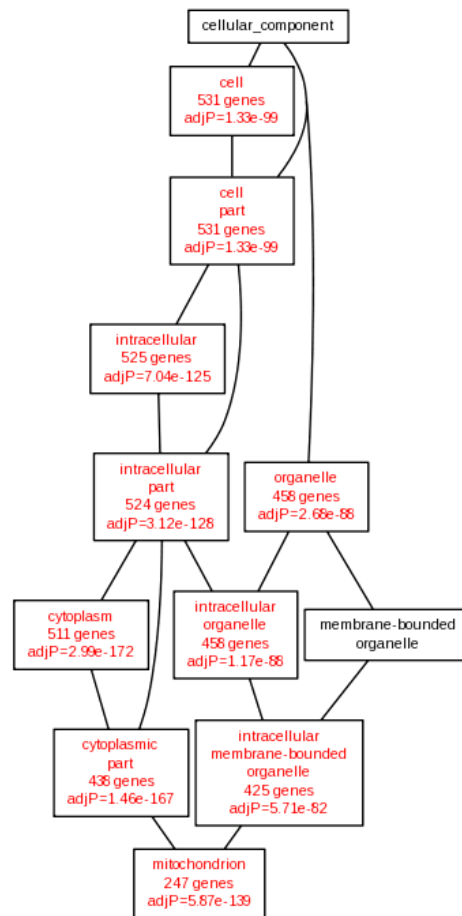


Figure 4.53.: Enrichment analysis of cellular component of 605 proteins identified in mouse liver within all biotinylation reagent groups and the negative controls (without one hit wonders). Visualisation as directed acyclic graph using Webgestalt.^{[283],[284]} Enriched GO categories are depicted in red, name of the GO category, number of included genes and adjusted p-value are shown.

The increased background proteome fraction in all samples reduces the detectable variability: The regulation between different sample groups is low and reactivity-dependent shifts are barely visible (see Figure 4.51). Proteins annotated to the plasma membrane category do not exhibit any statistically significant regulation between commercial or peptide-based reagents and the PBS-perfused negative control (unpaired t-test). This might be influenced by the non-specific co-purification of proteins with lipid-binding domains in all samples. Furthermore, the highly abundant background proteome might suppress the proper analysis of the labelled, but still low-abundant target proteome: The total number of identified proteins with plasma membrane or extracellular matrix annotation is not significantly increased compared to the negative controls (see Figure 4.47).

To summarize, it can be stated, that the background proteome fraction identified in the liver dataset severely interferes with the analysis: Differences among reagent groups get superimposed by non-labelled, co-identified proteins, also leading to the observed decrease in the surface proteome fraction. Hence, for a successful *in vivo* biotinylation experiment, not only the vascularisation of the target tissue plays a role: High lipid contents might interfere with sample preparation and hinder the generation of a valuable mass spectrometric dataset.

4.4.5.2. SRM-Analysis of High- and Low-Abundant Proteins

SRM analysis of candidate proteins was performed as described in chapter 4.4.4.3 with three biological replicates per group. Integrated information about protein abundance in mouse liver tissue was again extracted from PaxDB, based on several publicly available liver datasets.^{[280],[372],[371],[386],[387],[388]}

As shown before for mouse kidney tissue, the intracellular proteins Calreticulin, 60 kDa heat shock protein and Nucleolin also belong to the top 5% liver proteins in terms of protein abundance. The mitochondrial transmembrane proteins Sideroflexin-5 belongs to the top 25% of the liver proteome and is both a citrate and also potentially an iron transporter.^{[389],[390],[391],[392]} All of those highly abundant intracellular proteins could be stably quantified within the full proteome samples, and were down-regulated in the biotinylation approach (see Figure 4.54a-d).

The co-enriched biotin-carrying carboxylases are again heavily enriched by capturing on the streptavidin-sepharose. Acetyl-CoA carboxylase as well as Propionyl-CoA carboxylase belong to the top 5% in terms of protein abundance and could also be quantified out of the full proteome. Methylcrotonyl-CoA carboxylase is among the top 25%, but quantification out of the full proteome sample was already borderline (see Figure 4.54e-g). As already seen within the kidney SRM dataset, the top 25% line seems to be a good estimate about the measurability of protein candidates out of the full proteome lysates.

Voltage-dependent anion-selective channel protein 1 belongs to the top 10% of the mouse liver proteome in terms of protein abundance. The protein is reported to be localised in the outer mitochondrial membrane and is involved in the regulation of apoptosis.^[393] There is also high evidence that the protein is expressed

in the plasma membrane, functions are still intensively discussed.^[394] Not surprisingly for the high protein abundance, the protein was reliably measurable out of full cell lysates (see Figure 4.54h). The very stable quantification that was possible with the biotinylation samples might either emphasize the plasma membrane localization of the protein or indicate the co-purification of the high-abundant mitochondrial protein bound to lipid impurities in the samples subjected to digest with the streptavidin-sepharose-bound proteins. The latter hypothesis is supported by the highly hydrophobic nature of the protein forming a β -barrel composed of 19 strands within the membrane.^[394] Secondly, its enrichment is also observed in the PBS- or unperfused negative controls. Thirdly, variability between biological replicates is high, also suggesting differential amounts of co-purified lipid-bound proteins during sample preparation.

Finally, some proteins annotated to the plasma membrane or the extracellular matrix with medium to low abundance were examined (see Figure 4.54i-l). Again, the cell adhesion protein Cadherin-2 and the extracellular matrix component Collagen alpha-1(VI) chain were quantified. The latter belongs to the top 25% most abundant liver proteins and could be quantified out of full proteome lysates. Cadherin-2 as well as the intercellular desmosome junction component Desmocollin-2 and the plasma membrane localised or secreted Semaphorin-4G, interacting with Plexins especially in the nervous system, show medium to low protein abundances.^{[395],[396]} SRM analysis of those proteins was not possible with the full lysates. In accordance to the results from the kidney SRM analysis, the heparin-based reagents were down-regulated, whereas peptide-based and commercial reagents both enabled the SRM-based quantification of low-abundant surface or extracellular matrix annotated biomarker candidates.

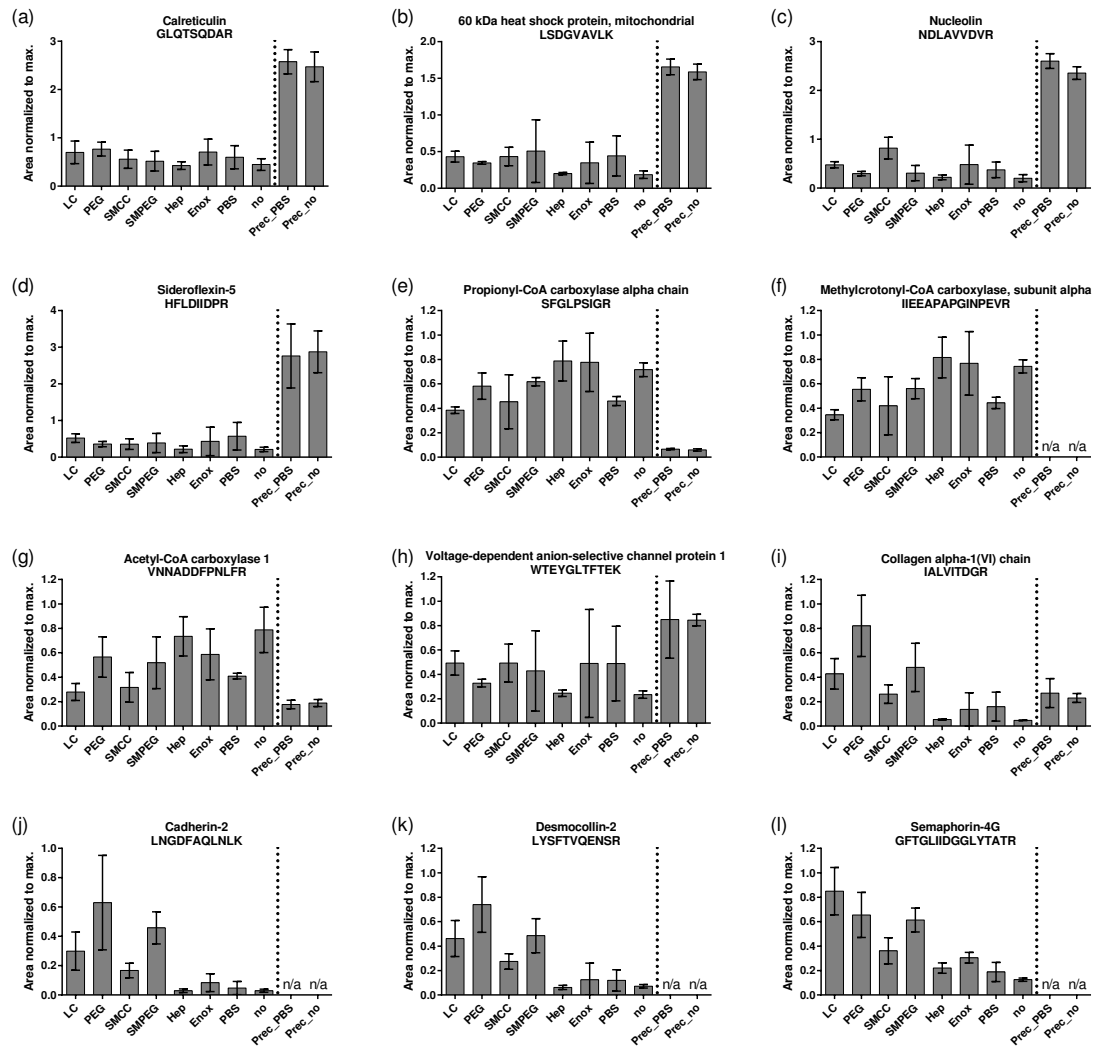


Figure 4.54.: SRM analysis of liver samples, mean with standard deviation of three biological replicates. Vascular accessible proteome enrichment with commercial reagents Sulfo-NHS-LC-biotin (LC) and NHS-PEG₁₂-biotin (PEG), peptide-based reagents SMCC-Cys-(L-Asp)₃-biotin (SMCC) and SM(PEG)₆-Cys-(L-Asp)₃-biotin (SMPEG), heparin-based reagents Biotin-Heparin-sNHS (Hep) and Biotin-Enoxaparin-sNHS (Enox), negative control PBS- (PBS) or non-perfused (no). Full proteome samples of precipitated proteins from PBS- (Prec_PBS) or non-perfused (Prec_no) mice. Proteins with intracellular annotation (a-d), carboxylases (e-g), cell membrane or extracellular matrix annotation (h-l).

5. Discussion

The aim of the work presented in this thesis was the development and validation of novel reagents for the enrichment of vascular accessible proteins, which is a prerequisite for the identification of novel disease-specific biomarkers, targetable for diagnostic or therapeutic purposes.

Potential biomarkers accessible for monoclonal antibodies are located on the cell surface, on newly formed blood vessels and in the perivascular extracellular matrix.^[160] Proteomic discovery approaches have to deal with the high sample complexity and a dynamic range of at least seven orders of magnitude in protein expression levels.^[41] Cell membrane proteins are highly diverse in function, but mostly expressed at low abundance levels, leading to their under-representation in full proteome datasets.^[132] Additional challenges in membrane proteomics are the decreased solubility of membrane proteins due to their hydrophobic character as well as the lack of polar cleavage sites for the standard proteolytic enzyme trypsin in trans-membrane domains.^{[133],[119]}

Classical fractionation techniques such as density gradient centrifugation cannot provide a clear separation of plasma membrane and intracellular membrane sub-proteomes (e.g. Golgi, endoplasmic reticulum, mitochondrion, nucleus).^[133] Several approaches have been developed to label the target proteins in the natural environment for subsequent enrichment: Carbohydrate-based cell surface capturing addresses glycoproteins via chemical modification of the sugar residues.^[172] Silica coating is based on vascular perfusion with colloidal silica followed by density-gradient-based enrichment of silica-coated membranes.^[174] The method of choice for covalent labelling of all proteins on the cell surface and within an accessible radius around the blood vessels in the perivascular matrix is the attachment of a derivative of biotin, subsequently used for streptavidin-based enrichment.^[160] The biotinylation technique is applicable *in vitro*, but also in *in vivo* disease models or *ex vivo* with surgically resected tissue material by vascular perfusion.^{[181],[184],[189],[190]}

The different studies performed with commercial biotinylation reagents have shown a substantial fraction of identified proteins with intracellular localization annotation, despite the enrichment of plasma membrane proteins. One reason is the plasma membrane crossing potential of the commonly used Sulfo-NHS-LC-biotin. In 2010, the novel, triply charged and size increased reagent NHS- β -Ala-(L-Asp)₃-biotin was published.^[12] The reagent was shown to up-regulate proteins with plasma membrane or extracellular matrix annotation compared to commercial Sulfo-NHS-LC-biotin. Nevertheless, only 219 surface or extracellular matrix annotated proteins could be identified in kidney tissue, originating in both the reduced reactivity of the reagent and the use of a meanwhile outdated MS system.^[12]

Within this thesis, reactivity tests on BSA as model protein revealed the poor reactivity of NHS- β -Ala-(L-Asp)₃-biotin in comparison to commercial reagents (see Figure 4.4). According to these results, two novel peptide-based reagents SMCC-Cys-(L-Asp)₃-biotin and SM(PEG)₆-Cys-(L-Asp)₃-biotin have been designed to improve the reactivity while maintaining features such as the increased size and the number of charges of the peptide linker. Additionally, two heparin-based reagents with a large-scale and highly negatively charged linker have been created. Synthesis was optimized for all compounds and the subsequent validation was performed against commercial reagents using systems with increasing complexity: (i) Reactivity was assessed on BSA as model protein. (ii) A proof-of-principle study was performed on HeLa cells. (iii) *In vivo* biotinylation efficacy was determined via perfusion of mice. Validation techniques included visualisation methods such as CLSM, FACS, Western Blot and ELISA as well as mass spectrometric analysis of the biotinylated proteome fraction of cells, mouse kidney and liver tissue enriched on streptavidin-sepharose beads.

5.1. Design of Novel Enrichment Reagents

The common characteristic of all enrichment reagents designed and used within this thesis is the activated carboxylic group for chemical derivatization of primary amino groups from lysine side chains or the N-terminus of the target proteins. To ensure the labelling of all accessible proteins, the targeted functional group has to be accessible for the reagent: Aliphatic amino acids as well as thiol groups from cysteine residues involved in disulfide bridges are therefore excluded. Secondly, the targeted functional group has to be reasonably abundant expressed: Lysine content in the total amino acid composition of the human proteome is about 5%.^{[294],[397]} Lastly, to minimize effects on cell viability and plasma membrane integrity, the labelling reaction has to be efficient and fast. The usage of activated carboxylic groups provides labelling in a one-step reaction and was shown in an activity test on BSA to react with amino groups within seconds (see Figure 4.3). Usage of NHS- or sNHS-activated carboxylic groups is therefore the method of choice for the application.

Activated carboxylic groups are prone to hydrolytic deactivation. Reactivity tests on BSA revealed the lowest hydrolysis rates at slightly acidic to physiological pH, perfectly matching with *in vitro* or *in vivo* applications carried out in PBS (see Figure 4.3). To prevent hydrolysis, the peptide-based reagents were synthesized in dry solvents and diluted directly before use. Although it was shown, that storage in dry DMSO under inert gas only leads to a slight reactivity decrease, all reagents used within this thesis were freshly synthesized to ensure comparability. Batch-to-batch variations detected by reactivity assessment on BSA most likely occur due to slightly different storage or freeze-thaw conditions leading to varying water content in the dried commercial reagents or in the peptide educt (see Figures 4.3 and 4.8). Moreover, different pH in the lyophilized peptide might occur, leading to different hydrolysis rates. Four peptide synthesis batches have been used within this

thesis. All individual batches were validated by reactivity assessment on BSA. For *in vivo* validation experiments two different peptide educt batches had to be used to synthesize the necessary reagent amounts. Heparin-based reagents are not soluble in organic solvents, therefore activation was performed in aqueous buffer directly before use.

The linker between the functional group for protein modification and the biotin residue for enrichment is both a spacer minimizing steric hindrance during labelling and capturing and determines the properties of the reagent.

The linkers of the novel peptide-based reagents SMCC-Cys-(L-Asp)₃-biotin and SM(PEG)₆-Cys-(L-Asp)₃-biotin exhibit an increase in charge and molecular weight compared to the commonly used commercial Sulfo-NHS-LC-biotin with an aliphatic linker. Those characteristics not only increase water solubility, but also minimize the membrane crossing potential of the reagent. In contrast to the biotinylation reagent NHS- β -Ala-(L-Asp)₃-biotin published by Strassberger *et al.*, the novel reagents can be activated in a site-specific manner by a bifunctional crosslinker targeting the reagent's C-terminal cysteine residue, thus minimizing steric hindrance during the labelling reaction on proteins. Hereby, reactivity was more than doubled in the activity test on BSA as well as in FACS analysis on biotinylated HeLa cells (see Figures 4.7, 4.14), while the linker structure with all its positive properties is kept.

Additionally to the peptide-based reagents, heparin chains were chosen as second linker type (Biotin-Heparin-sNHS and low molecular weight Biotin-Enoxaparin-sNHS). The large size, the high polarity with multiple charges, the extreme hydrophilicity and the presence of a variety of functional groups for chemical derivatization make the heparin structure a promising linker type. The reducing chain ends were site-specifically derivatized with biotin hydrazide (exactly one biotin residue per molecule) enabling direct comparisons of the biotin signal with the commercial reagents via CLSM analysis of tissue sections or ELISA. Heparin chains contain one carboxylic group per disaccharide subunit, which can be activated for protein coupling. FACS analysis of labelled HeLa cells has shown that full NHS activation of all carboxylic groups provides the best labelling rates possible. Chemical analysis of the heparin-based reagents as well as reactivity assessment on BSA is complicated due to its polymer nature with various heparin chain lengths and its extremely high mass. Therefore, linear mass spectrometric analysis of labelled BSA to prove successful labelling was only possible when using low molecular weight enoxaparin in a reagent to protein ratio of 1:1 (see Figure 4.10). Successful activation of the carboxylic groups could additionally be shown in a mass spectrometric experiment of labelled HeLa cells: Non-activated Biotin-Heparin yields a result comparable to the negative control, while proteins are efficiently labelled and stably identified upon NHS- or sNHS-activation of the heparin chains. Biotin signal could be proven by CLSM analysis and ELISA of biotinylated tissue samples.

The biotin-streptavidin system is used for enrichment, as the interaction is one of the strongest non-covalent interactions known ($K_d \sim 10^{-15}$ M)^[291], providing a stable platform for affinity-based purification. As the captured proteins are directly digested on the beads, also a covalent and bioorthogonal capturing strategy could

in principle be developed. Circumventing the biotin system would also provide the chance to avoid the co-purification of high-abundant intracellular proteins intrinsically carrying Biotin, such as carboxylases. Click chemistry capturing based on an azide-alkyne system is such a bioorthogonal method. Therefore two alkyne-based reagents based on the triply negatively charged peptide linker also used for the biotinylation reagents have been synthesized. Reagents with terminal alkynes have to be captured on azide-coupled beads under Cu(I) catalysis. As the labelling reaction on cells or via the *in vivo* perfusion is still based on the reaction of the reagents' activated carboxylic group with a primary amino group, copper is hereby not applied in living species, but in protein lysates. Nevertheless, some combinations of copper and Cu(I) stabilizers have been found to favor protein precipitation. That's way additionally a reagent based on a dibenzylcyclooctyne was synthesized, a system that is highly reactive due to the enormous ring strain and not copper-dependent. As expected, reactivity of the alkyne-based reagents is comparable to the peptide-based biotinylation reagents. Thiol-yne click coupling reported to occur as side-reaction between thiol groups and alkyne residues in some systems seem to play a minor role for this application.^{[398],[399]} Click-mediated capturing on azide-modified beads is currently set up. The low concentration of the modified proteins within the lysates is one of the challenges for the approach.

5.2. Reactivity and Visualisation of Biotinylation

The site-specific activation of the novel peptide-based biotinylation reagents heavily improves the reactivity compared to non-specifically activated NHS- β -Ala-(L-Asp)₃-biotin published by Strassberger *et al.* as shown by reactivity assessment on BSA (see Figure 4.7).^[12] Nevertheless it is not possible to achieve the same reactivity as with the commercial Sulfo-NHS-LC-biotin: The peptide linker chosen to introduce charges and an increase in size results in a structure sterically more hindered in comparison to the aliphatic Sulfo-NHS-LC-biotin linker.

Besides the reactivity assessment on BSA, differential reactivity could also be visualised by FACS analysis of HeLa cells labelled in different ratios with the reagents (see Figure 4.15). CLSM analysis provided spatial biotinylation information: Plasma membrane labelling could be visualized with both commercial and peptide-based reagents (see Figure 4.16). Hereby, the decreased reactivity of the novel peptide-based biotinylation reagents was obvious, as the total signal intensity was decreased. Consequently, samples biotinylated with peptide-based reagents had to be acquired with higher laser intensities than the commercial reagents to ensure consistent quality of the pictures. PBS-treated cell were taken as negative controls in both cases. Based on the experience with the capacity of the LC/MS system, the percentage of biotinylated proteins is about 1-2% of the total protein content of the cell lysates. Analysis of biotinylation via Western blot was difficult: Vast total protein amounts had to be loaded to get a biotin signal over the whole protein lane. Without severe restriction of the quality of the Western Blot, this approach could

only be performed with the higher reactive commercial reagents.

Within tissue samples of perfused mice it could be demonstrated that all biotinylation reagents are capable to label accessible proteins within a few cell layers around blood vessels and to enter highly capillarized regions, which nicely corresponds to the tissue parts that can be targeted by a monoclonal antibody vascularly administered for diagnosis or therapy (see Figure 4.30). With CLSM analysis, no significant differences between commercial and peptide-based reagents could be detected. Within liver tissue, staining mainly occurred around vascular structures, whereas in kidney tissue the overall staining was stronger with an additional accumulation of the reagent around tubular structures and in the glomeruli. Stronger kidney biotinylation was also proven by ELISA performed with the highly concentrated kidney and liver homogenates. Western Blot analysis was only possible with the higher biotinylated kidney lysates. Similar to the *in vitro* analysis, Western Blot lanes have to be overloaded to get signals from the peptide-based reagents (see Figure 4.34). Nonetheless, there is no visible difference between commercial and peptide-based reagents in the CLSM analysis (see Figure 4.31). Within ELISA analysis, SMCC-Cys-(L-Asp)₃-biotin turned out to be slightly more reactive than the sterically more hindered SM(PEG)₆-Cys-(L-Asp)₃-biotin (see Figure 4.33), which was also confirmed in the mass spectrometric analyses of cells or tissues.

The reactivity of heparin-based reagents was decreased compared to the commercial and the peptide-based biotinylation reagents as shown in all analyses. FACS analysis of HeLa cells biotinylated with different excess and differential NHS activation grade of the heparin-based reagents showed the best results, when every disaccharide unit contained an activated carboxylic group (see Figure 4.17). On cells, no biotin signals could be detected via CLSM analysis or Western blot. CLSM analysis of tissue sections of perfused mice revealed a comparable biotinylation pattern to the commercial and peptide-based reagents, while overall signal intensity was significantly decreased (see Figure 4.32). Therefore, pictures had to be taken with higher laser intensity to ensure consistent quality preventing a direct comparison of the signal intensities across different reagent groups. PBS- and non-perfused mice were taken as corresponding negative controls, pictures were hereby acquired with all different laser properties used within the analyses. ELISA-based analysis of the biotinylation rate could detect significantly increased biotin content in comparison to the PBS-perfused negative control (see Figure 4.33). Direct quantitative comparison to the commercial reagents is not possible within the same experiment to ensure a linear correlation of the ELISA biotin standard in the acquired range. Western blot analysis was not possible on tissue homogenates biotinylated with the heparin-based reagents.

Low biotin signals in all streptavidin detection-based tests might point towards a decreased reactivity of the heparin-based reagents, but also leaves room for an additional explanation: Due to the enormous tag size of the heparin-based linkers, less biotinylated reagent molecules can be attached per surface area in comparison to the small commercial or peptide-based reagents. This is why the maximal log-shift achievable by FACS analysis of biotinylated HeLa cells is about two log-shifts less

than for commercial Sulfo-NHS-LC-biotin (see Figure 4.18). Hereby, no significant difference between the heparin and the low molecular weight enoxaparin linker could be detected. Mass spectrometric analysis of kidney tissue nevertheless revealed a reactivity decrease of the heparin-based reagents.

For mass spectrometric-based validation, all cell samples or mice were biotinylated with the same amount of biotinylation reagent allowing for a comparable analysis. For the *in vitro* experiment, the amount was determined by FACS analysis, so that maximal log-shift was achieved for all reagents. *In vivo* biotinylation was performed with an amount yielding to comparable biotin signal intensities of commercial and peptide-based reagents in CLSM analysis of tissue sections. Mass spectrometric analysis of *in vitro* samples still revealed a down-regulation in signal intensity with the peptide-based reagents corresponding to the decreased reactivity, whereas the *in vivo* analysis was equivalent across the peptide-based and commercial reagent groups. So, CLSM analysis might be the best way to evaluate optimal reagent amounts for a valuable mass spectrometric analysis.

5.3. Aspects of Sample Preparation for Mass Spectrometric Analysis

Differential FACS analysis of biotinylated HeLa cells for reactivity assessment was performed on detached cells to ensure constant cell counts. Hereby, non-tryptic detachment was achieved with 10 mM EDTA to preserve the cell surface proteins. The slow detachment procedure decreases cell viability to below 80%. The consequence is the biotinylation and mass spectrometric identification of a large fraction of intracellular proteins (up to 85%) most likely derived from membrane permeable, necrotic or apoptotic cells. Therefore, for successful mass spectrometric analysis of the cell surface proteome, cells have to be biotinylated following a washing protocol, but still adherent to the cell culture flask to ensure maximal viability. As the biotinylation solution has to be applied in lower concentration on adherent cells in comparison to biotinylation in solution, the total amount of biotinylation reagent had to be increased about ten-fold to 1 μmol reagent per 1×10^6 cells. Hereby, optimization was performed via FACS analysis to achieve maximal log-shift with all reagents. Cells were seeded in cell culture flasks 24 h prior to biotinylation in constant cell density to ensure stable cell counts for the validation experiments. Non-reacted reagent was quenched with amine-containing buffer before proceeding with washing, detachment and cell lysis, so that no further labelling non-specifically occurred in the following workflow.

The novel peptide-based biotinylation reagents were synthesized in dry DMSO to prevent hydrolysis. DMSO is one of the most polar organic solvents ensuring the solubility of the peptide-based reagents in high concentrations, but difficult to remove due to its high boiling point. The biotinylation reagent solution was diluted in PBS before use resulting in a final DMSO concentration below 1%. No influence

on the cell viability could be detected at the concentrations used.

Healthy NSG mice were perfused for *in vivo* validation of the novel biotinylation reagents. All perfusion solutions contained 10% Dextran-40 000 MW to increase the oncotic pressure and subsequently increase the flow within the blood vessels while minimizing the risk of thrombosis, which would inhibit the flow through the blood vessels.^[185] Perfusion solutions and mice were kept at body temperature for maximal perfusability. During perfusion, blood components were washed out. To ensure the inactivation of non-reacted reagent, mice were sprinkled with amine containing solution during the procedure and perfusion with the reagent was followed by perfusion with amine-containing quenching solution. CLSM and ELISA analyses have shown differences in the degree of biotinylation between different organs, different reagents as well as among biological replicates: Biotinylation success is dependent of the tissue vascularisation, the reagent's reactivity, reagent's batch-to-batch variations, as well as of the technical quality of the perfusion.

To solubilize membrane proteins from the surrounding lipid layers, different detergents such as SDS and sodium deoxycholate were used in the solubilisation buffers in about 2% total concentration. In such concentrations, these detergents are known to negatively interfere with enzymatic digest and mass spectrometric analysis, which led to the development of cleavable surfactants or other strategies such as gel filtration in 8 M urea for full proteome samples.^{[120],[135]} Due to the capturing of the biotinylated proteins on the streptavidin resin followed by separation from the filtrate and extensive washing, the detergent containing protein extraction buffers are compatible with the LC/MS analysis of the enriched surface proteins. Full proteome samples have been prepared for SRM analysis out of the protein lysates from the PBS- or non-perfused negative controls also used for the surface proteome enriched sample, by using RapiGest™ (waters), an acid labile surfactant that is separated before peptide desalting by centrifugation following cleavage of tridecan-2-on.

Reduction of disulfide bridges followed by alkylation of the free thiol residues with iodoacetamide helps protein unfolding and prevents reformation of the disulfide bridges. Thus, tryptic digest and consequently sequence coverage are improved.^{[348],[349]} Capturing of the biotinylated proteins was carried out under reductive conditions using TCEP, because TCEP is reported to be more efficient than DTT and cannot react with the alkylating agent.^{[349],[400]} As the proteins are bound to the streptavidin resin, removal of excess alkylating agent in solution could be easily achieved to prevent over-alkylation or non-specific alkylation of other amino acid side chains by long incubation times.^[401] Delipidation of full proteome samples is commonly carried out by chloroform/methanol- or acetone-based protein precipitation, while lipids are kept in solution.^{[281],[402]} For the tissue lysates prepared, delipidation turned out to be crucial for efficient sample preparation and valuable mass spectrometric analysis. Since the target proteins are bound to the resin, a precipitation-free method based on extraction in organic solvents (1-butanol/diisopropylether solution) had to be applied.^[347]

Tryptic digest is carried out directly on the streptavidin-sepharose bound proteins. The modified trypsin (Promega) used provides high specificity as it was

treated with TPCK, a chymotrypsin inhibitor: Autolytic products of native trypsin have chymotrypsin-like activity, leading to the generation of peptide fragments with non-tryptic cleavage sites interfering with database search.^[403] Furthermore, the enzyme used contains methylated lysine residues to provide stability against autolytic digestion.

The mass spectrometric detection of peptides is severely influenced by the presence of nonvolatile salt contaminants in the analyte: Decreased ionisation efficiency, ion suppression, adduct formation and increase of noise in the spectra is the consequence.^{[404],[405]} Therefore, sample desalting is essential for mass spectrometric analysis, despite of a certain peptide loss: Most desalting methods are reported to have 98-99% desalting efficiency with peptide recovery rates above 70%, peptide binding can hereby be increased by aspirating the sample in several cycles.^[406] Within this thesis, desalting tips as well as desalting plates with C18 material were applied for sample desalting, material was loaded with respect to the reported binding capacities by twenty aspiration cycles (desalting tips) or in three loading steps (desalting plate).

The peptide mixtures obtained from enzymatic digestion from full proteome or plasma membrane proteome enriched samples are highly complex (diversity), and highly variable in abundance (dynamic expression range). Powerful complexity reduction steps such as reversed-phase capillary liquid chromatography-based separation prior to mass spectrometric analysis are crucial to obtain valuable datasets.^[407] Small particles of 1.4-3 μm size in the packed column medium are reported to increase the peak capacity during chromatographic separation, but the operating pressure also has to be raised: The nanoACQUITY system (Waters) used is operated with a C18 column with 1.7 μm particle size and 10 000 psi pressure limit.^{[200],[408]}

For the discovery experiments, a MALDI/TOF system was used for analysis. Chromatographic fractionation, MS and MS/MS analysis are uncoupled (off-line process) providing the possibility to archive the sample plates and to immediately evaluate sample quality after MS analysis (2D peptide maps). Compared to an ESI system, analysis speed is decreased, but no peptide component is missed during MS analysis due to measurement time aspects and MS/MS data is stably collected from the maxima of the chromatographic elution peaks. Furthermore, MALDI is described to be less sensitive to ion suppression by interfering contaminants.^[409] Both ionisation by ESI and MALDI provide accurate data, the methods are reported to be complementary with a significant overlap of above 60% in protein identification.^{[410],[409]} On peptide level, MALDI tends to an increased identification of larger peptides with higher retention times on reversed-phase columns.^{[409],[411]} With the off-line system used, stable, label-free, relative quantification is enabled, as several standard peptides covering the mass range were spiked into every fraction for spectra normalisation.

5.4. Presented Datasets and Aspects of Data Analysis

Mass spectrometric validation of the novel peptide- and heparin-based biotinylation reagents was performed against the commercial reagents Sulfo-NHS-LC-biotin and NHS-PEG₁₂-biotin. PBS-treated cells or PBS- and non-perfused mice were taken as negative controls to investigate non-specific protein identification.

As shown previously, different biotinylation reagents have different structures, characteristics and reactivities, leading to different amounts and also types of labelled proteins. Therefore, the amount of biotinylated protein in the lysates is not a good basis for a comparative analysis. Furthermore, analysis of the biotin content by ELISA would have to deal with technical challenges such as low total protein concentrations in the cell lysates, differences in signal intensities inhibiting the analysis of all samples within the same linear standard range as well as with the excess of activated carboxylic groups in the heparin-based reagents possibly leading to the labelling of several proteins on the same singly-biotinylated heparin molecule. Thus, the comparative analysis is performed based on biotinylation with the same reagent amounts (5 μ mol per cell culture flask with about 5×10^6 cells and 25 μ mol per mouse) followed by subjecting the same total protein amount to capturing on the streptavidin-sepharose beads. The amount analysed per LC/MS runs always refers to the total protein amount, which the captured and digested peptides were derived from (discovery experiments: *in vitro*: 500 μ g, kidney: 750 μ g, liver: 1.25 mg). The amounts are chosen to use the full capacity of the LC/MS system with the highest biotinylated samples, which could be easily visualized by 2D peptide map densities of the MS1 runs. As the biotinylation rate of liver tissue is lower than for kidney tissue (shown by CLSM analysis and ELISA), the total protein amount subjected to analysis was higher than for kidney tissue.

In the following, the structure and the analysis strategy of the datasets presented in this thesis are elucidated. For *in vitro* validation of the peptide-based reagents, technical triplicates of HeLa cells biotinylated with SMCC-Cys-(L-Asp)₃-biotin, SM(PEG)₆-Cys-(L-Asp)₃-biotin, Sulfo-NHS-LC-biotin, NHS-PEG₁₂-biotin or treated with PBS were analysed. The heparin-based reagents Biotin-Heparin-sNHS and Biotin-Enoxaparin-sNHS were analysed against Sulfo-NHS-LC-biotin and PBS treatment in technical duplicates. *In vivo* validation was performed via perfusion of mice with Sulfo-NHS-LC-biotin, NHS-PEG₁₂-biotin, SMCC-Cys-(L-Asp)₃-biotin, SM(PEG)₆-Cys-(L-Asp)₃-biotin, Biotin-Heparin-sNHS, Biotin-Enoxaparin-sNHS, PBS or non-perfused as negative controls and analysis of kidney tissue with four biological replicates per sample in technical duplicates each. As the variability between technical replicates in kidney tissue could be shown to be low, liver tissue was analysed in groups of biological triplicates only.

The proteotypicity of the identified peptides was determined with respect to a given protein database of the organism. Only a proteotypic peptide can be clearly mapped to its protein of origin, degenerated peptide would lead to ambiguities in

protein identification and also interfere with protein quantification.^[412] There are several possibilities to deal with non-proteotypic peptides in data analysis. The one applied for data analysis within this thesis is the most unambiguous one: All non-proteotypic peptides as well as peptides identified below the 95% confidence filter limit were excluded from analysis.^[413] The disadvantage of the applied strategy is the possible loss of valuable data. Another approach reported in literature is the grouping of degenerated peptides based on the amino acid sequence to protein families (termed peptide-sharing closure groups) and the analysis of abundance changes of whole protein families, which often share sequences as well as biological functions.^[414] Although all data is conserved, it is challenging to analyse differences between proteins that share many peptides.^[415] The third and most elaborated possibility is the assignment of the degenerated peptide entries to specific proteins based on abundance criteria such as spectral counting determined with the non-degenerated peptides.^[416]

To evaluate differences in the identification of specific proteins with different biotinylation reagents, the total number of identified proteins as well as the numbers of proteotypic peptides identified per specific protein within a sample were analysed. To obtain a stable dataset and to compensate for technical run-to-run variability or variability among biological replicates, the information from all replicate runs belonging to one reagent group were hereby summed for analysis. For further analysis of the datasets, so-called one hit wonders were excluded if not indicated differently: Proteins analysed were identified with more than one proteotypic peptide in at least one sample group.

Relative quantification on the validation datasets was performed based on MS1 signal intensities using the in-house developed software MS_QBAT.^[282] A fraction-wise normalisation against four internal spike-in peptides in every MS1 spectrum compensated for run-to-run differences. A more stable quantification can be achieved by reduction of noise via exclusion of features below a given S/N filter and feature length: Filter criteria were chosen in order to keep 99.5% of all identifications, while about 40 - 50% of all features could be dismissed. Annotations were transferred between technical and biological replicates under usage of narrow alignment parameters for fraction and mass delta to minimize the risk of false alignments.

Proteomic datasets are reported to exhibit a median overlap between technical replicates of 75% in terms of protein identifications and 35 - 60% in terms of peptide identifications.^[417] With the data analysis strategies chosen it was possible to obtain stable, valuable datasets: Main points hereby are the propagation of annotations between technical and biological replicates as well as the combination of replicates to super-samples. Quantification of all pairs of technical replicates within the *in vitro* dataset shows a median regulation of 1.05 with more than 90% of all values below a two-fold regulation (see Figure 4.25). Variability of technical replicates in kidney samples was also shown to be low, median regulation across eight sample pairs was 1.01 with 90% of all values below a 1.57-fold regulation. Biological replicates exhibit a broader variability across eight exemplary sample pairs, but still 50% of the values were below an 1.3-fold regulation and 90% below a 2.2-fold regulation (see

Figure 4.40). Differences in biotinylation efficacy within the same reagent group can lead to a slight up-regulation of the ratio towards the stronger biological replicates. To compensate differences derived from technical variability in the biotinylation workflow and the subsequent mass spectrometric analysis, all samples belonging to the same reagent group were combined to super-samples via feature alignment by an iterative process (genetic algorithm) based on a given fraction and m/z range and the present annotations. Those super-samples were finally quantified against the super-sample of the commercial Sulfo-NHS-LC-biotin reagent group. In a presence vs. absence situation (e.g. the quantification of most of the proteins between Sulfo-NHS-LC-biotin and the PBS-treated *in vitro* negative control), artificial background values were introduced to enable relative quantification. Similar to the analysis of the summaries of proteotypic peptides, variability was reduced if proteins quantified with only one proteotypic peptide were excluded.

To further elucidate the nuclear protein enrichment with the heparin-based reagents *in vitro* as well as the identification of the background proteome in tissue samples, additional datasets were generated based on single samples prepared with different strategies. Hereby, all samples biotinylated with the same reagent within an analysis were derived from the same cell lysate or the same tissue sample to reduce biological variability. Whereas samples of the *in vitro* or *in vivo* validation experiment described before were analysed consecutively, without any avoidable instrumental variability, the background proteome investigation was performed in several steps, over a long period of time, with large mass spectrometric performance differences (determined by repeated measurements of an *E. coli* digest standard). To account for those differences, MS_QBAT was used to propagate annotations between runs of sample within the same reagent group: Features from MS1 data were extracted and cross-alignments between all possible sample pairs were performed with narrow alignment parameters for fraction and mass delta to minimize false positive alignments. Annotation information was transferred, if no annotation of that feature was present in one sample, but in the other. Data analysis was performed based on identified protein numbers and their localization annotation.

To analyse the subcellular localization of the identified proteins, the annotation information was extracted from the UniProt database. The information was manually curated by screening for different categories like plasma membrane, extracellular, intracellular and membrane associated. As plasma membrane enrichment was used, proteins exhibiting both plasma membrane and intracellular localization annotations were assigned to the plasma membrane category. Proteins annotated with not further specified membrane association (contain hydrophobic domains) are most likely also associated with the plasma membrane, but are kept in an own category. In some exceptional cases, the manual analysis might be ambiguous. Nevertheless, it should be considered that biological systems are complex, for example proteins may translocate upon different conditions or in malignant cells, leading to several contradictory or incomplete annotations in the database.^{[16],[17],[18],[19]} Still, automated enrichment analyses of protein subsets using Webgestalt, supported the manual localization assignment and are also presented within this thesis.

5.5. The Background Proteome

The aim of the biotinylation technique is the stable identification and quantification of plasma membrane proteins *in vitro* or of vascular accessible proteins on the surface of cells, on newly formed blood vessels or in the perivascular extracellular matrix in an *in vivo* setting. Enrichment of the highly diverse and often low-abundant plasma membrane proteome fraction is successful with the peptide-based reagents as it was shown by a comparison of the *in vitro* data to a full proteome dataset by Nagaraj *et al.* as well as by SRM analysis of full proteome samples. In full proteome samples, in particular proteins with low or medium high abundance are either unstably identified (i.e. low numbers of proteotypic peptides in the discovery dataset or signals within the background in SRM analysis) or not found at all.^[130] In contrast, identification of high-abundant intracellular proteins such as proteasome subunits was significantly decreased within the surface proteome enriched samples. Nevertheless, the fraction of intracellular proteins identified within the surface proteome enriched samples is still about 50 - 60% within the validation on HeLa cells as well as on kidney tissue (see Figures 4.23, 4.35). Also in published datasets using commercial biotinylation reagents, high intracellular fractions to up to 70% are reported, but even if the high portions are mentioned, reasons are not further investigated in these biomarker studies.^{[180],[186],[183],[295],[296]} Within this thesis, the proteome fraction identified in the negative controls as well as different sample preparation strategies were further examined to elucidate reasons for the co-identification of intracellular proteins.

Only a small number of proteins was identified within the PBS-treated negative control in the *in vitro* dataset, namely 65 proteins out of a triplicate analysis. Top hits are different carboxylases carrying biotin as a cofactor, leading to their co-enrichment on the streptavidin-sepharose. Those enzymes were also found within the *in vivo* analysis and their biotin signal could additionally be detected via Western Blot of the negative control protein lysates from kidney tissue.

As the protein fraction within the negative control is so low, non-specific binding to the streptavidin-sepharose cannot be the main factor to explain the identification of hundreds of intracellular proteins within the surface proteome enriched samples. However, not all intracellular proteins identified have necessarily to be contaminations. The analysis of localization based on database annotation entries was already mentioned in the previous chapter: Entries might be not complete, as many examples for protein translocation from the intracellular part to the plasma membrane are found under specific conditions such as in a pre-apoptotic state or in cancer cell lines.^{[16],[17],[18],[19],[20]} Plasma membrane crossing by biotin transporters or passive diffusion due to the high concentrations on the cells is unlikely with the novel reagents, which are charged and increased in size, but might occur with the commercial Sulfo-NHS-LC-biotin having a small and non-polar linker.^[12] Co-purification of intracellular proteins attached to plasma membrane proteins by non-covalent ionic interactions can only play a minor role, as different washing strategies such as high salt buffer did not significantly influence the outcome. Enrichment analysis of the intracellular background proteome revealed the preferential identification of high-

abundant intracellular proteins such as ribosomal subunits. Those proteins are most likely identified due to covalent labelling with the biotinylation reagents. Thus, the undesired labelling must occur during the biotinylation procedure due to a small fraction of dying cells within the cell culture flask exhibiting permeable cell membranes or releasing proteins. As the cells are still adherent during the biotinylation procedure, cell viability could only be assessed after subsequent detachment and was determined to be below 80% at that time, also within the negative control.

Within the *in vitro* samples biotinylated with the heparin-based reagents, the intracellular fraction is increased to about 75% with a specific enrichment of the nuclear proteome (see Figures 4.27, 4.28). Besides the considerations discussed above, an additional factor not just depending on pure protein abundance must play a role. Reason for the nuclear proteome enrichment can be found in the structural similarity of the heparin chains to the DNA backbone.^{[350],[359]} Two hypotheses based on covalent binding of nuclear proteins from necrotic or apoptotic cells mediated by covalent attachment or non-covalent interactions of nuclear proteins to the heparin chains out of the cell lysate during sample preparation have been further discussed in chapter 4.3.4.2. Although the mechanism of nuclear protein co-enrichment could not yet be fully elucidated, it can be stated that the desired surface proteome enrichment with the heparin-based reagents is heavily suffering from the increased background proteome.

The fraction of intracellular proteins identified in the *in vivo* samples is comparable to the *in vitro* samples with about 50 - 55% in the kidney samples and 63 - 69% in the liver samples with peptide-based and commercial reagents. The fraction is slightly increased to about 60% (kidney) and 72% (liver) with the heparin-based reagent samples. Hereby, no significant nuclear proteome enrichment could be detected in the *in vivo* setting. This finding might support the hypothesis of covalent co-enrichment of nuclear proteins in the *in vitro* setting. Nevertheless, the main reason is the reduced reactivity of the heparin-based reagents artificially increasing the fraction of the background proteome detected within all samples.

In contrast to the *in vitro* negative control setting, hundreds of proteins were identified within the PBS- or non-perfused kidney and liver tissue. *In vivo* lysates are highly complex and larger total proteins amounts have to be subjected to capturing and the sample preparation as well as on the LC/MS system: The protein amount loaded on the same volume of streptavidin-sepharose is 2.5-fold (kidney) and 5-fold (liver) increased compared to the *in vitro* setting. Furthermore, the lysates are highly concentrated to enable comfortable sample handling with minimal streptavidin-sepharose losses. Thus, more unspecific binding to the resin might occur. However, introduction of washing steps with high salt buffer or pH gradients did not significantly change the result. Unspecific binding to the streptavidin-sepharose cannot explain the extent of the background proteome identified within the negative control samples. The most likely explanation is found within sample preparation. The desired proteome fraction is bound to the streptavidin beads and separated from the lysates by filtration. A non-soluble contamination within the tissue homogenates would be co-purified with the resin and co-subjected to tryptic

digest. To exclude the presence of solid particles or precipitated protein in the highly concentrated tissue lysates, ultracentrifugation followed by capturing of proteins out of the supernatant was examined, but did not change the result. Experimental observation of lipid structures within the homogenates blocking the filter tubes led to the introduction of an essential delipidation step during sample preparation. Hereby, the beads are incubated with a mixture of polar and non-polar solvents. Vast lipid amounts are visible at the interphase between the solvents. Lipid removal works well enough to not further interfere with the sample preparation, nevertheless, it is barely possible to separate the beads, solvents and the lipid interphase perfectly. If some of those non-soluble lipid structures are subjected to tryptic digest with the streptavidin beads, all proteins that bind to the lipid structures via hydrophobic domains would be co-digested. This hypothesis is supported by analysis of the proteins identified within the negative controls as well as within the core-proteome identified with all biotinylation reagents and the negative controls: The enrichment of proteins annotated to membrane bounded subcellular localization is striking. First of all, high-abundant metabolic enzymes localised in the mitochondrial membrane are identified. Although the mostly low-abundant plasma membrane fraction is significantly down-regulated in the negative controls in terms of numbers of identified proteotypic peptides as well as in quantified signal intensity in the kidney samples, it is noticeable that the plasma membrane fraction identified is still surprisingly high (up to 18% in kidney tissue), also supporting the hypothesis of lipid-bound protein co-purification and co-digestion. The use of another type of streptavidin beads such as magnetic beads might enable a more successful separation.

As the background proteome is high and both identified in the negative controls and the biotinylated samples, the variability and the visible reactivity shift detected by relative protein quantification across the samples is artificially decreased in the *in vivo* setting compared to the *in vitro* setting. The comparison of the kidney and the liver dataset reveals that the differences are even smaller among the liver samples: The overlap in protein identification between different sample groups is high (about 80%) and no clear dependence of the reagent's linker-type is visible. Relative quantification between different reagent groups results in a very narrow ratio distribution featuring similar characteristics as the quantification of pairs of biological replicates within the kidney dataset. Pronounced reactivity differences (down-regulation of the median) are barely visible and the similarity of the PBS-perfused and especially the sample biotinylated with NHS-PEG₁₂-biotin is obvious. The latter is probable an effect of the small dataset: Within the kidney dataset, eight samples (four biological replicates in two technical replicates each) were combined to super-samples, whereas the liver groups are based on three biological replicates, as the variability between technical replicates was low within the kidney set. The smaller dataset gives more room to outliers, in this case either biological replicates with higher lipid content and/or a less successful separation of lipid contaminants during sample preparation. These findings correspond to the experimental observation that liver samples contained more lipids than the kidney samples, complicating the sample preparation and the LC/MS-based analysis. Furthermore, the liver samples are less biotinylated

than the kidney samples as also shown by ELISA and CLSM analysis. Therefore, the total protein amount subjected to analysis is higher and the tissue homogenates even more concentrated than the kidney samples. The higher background may also explain the increased fraction of intracellular proteins within the liver dataset.

In particular in the liver dataset, the identified background proteome fraction severely interferes with the analysis. In summary, it can be stated that for a successful biomarker identification experiment, not only the often reported tissue vascularisation enabling the transport of the reagents to the site of interest plays a role, but also an elaborated delipidation strategy adapted to the lipid content of the respective tissue. Moreover, the preparation of PBS-perfused negative controls in biomarker studies helps to evaluate, if proteins are accessible by the vasculature and enriched or only co-purified.

5.6. Biotinylation Reagents: Reactivity vs. Selectivity

The *in vitro* dataset mirrors the reactivity differences among the commercial and peptide-based reagents detected by CLSM and FACS analysis: Relative quantification of all proteins against commercial Sulfo-NHS-LC-biotin revealed a two-fold down-regulation of the median with SMCC-Cys-(L-Asp)₃-biotin and a five-fold down-regulation with SM(PEG)₆-Cys-(L-Asp)₃-biotin, which is increased in size and sterically more hindered on the cell surface. 48% of the cell surface proteins (without one hit wonders) were identified with all biotinylation reagents. The less reactive reagent SM(PEG)₆-Cys-(L-Asp)₃-biotin does not stably identify a unique protein fraction, whereas the uniquely identified fraction is 7% in Sulfo-NHS-LC-biotin and 6% in SMCC-Cys-(L-Asp)₃-biotin. In accordance with the reactivity, the total number of identified plasma membrane proteins is slightly reduced with SMCC-Cys-(L-Asp)₃-biotin compared to Sulfo-NHS-LC-biotin. The reactivity differences result in decreased biotinylated protein amounts and subsequently also decreased signal intensities in LC/MS analyses, preventing the detection of larger variations in protein accessibility with the different reagents. As long as the cell viability is not affected, an increase in peptide-based reagent amount for further experiments is discussable. Nevertheless, the proof-of-principle study could show, that the novel peptide-based reagents are capable to enrich and stably identify and quantify proteins annotated to the cell surface.

Within the previous chapter it has been discussed that the ability of the large and multiply charged heparin-based reagents for cell surface proteome enrichment is suffering from the co-enrichment of nuclear proteins in the *in vitro* setting. The analysis of biotinylated kidney and liver tissue first of all revealed the decreased reactivity of the heparin-based reagents: Down-regulation of the median of all proteins quantified against commercial Sulfo-NHS-LC-biotin is 2-fold in kidney tissue. Comparison to the negative control down-regulation reveals the influence of the identified background proteome: The PBS-perfused negative control exhibits a 4-fold, and the more complex non-perfused control a 7-fold down-regulation, whereas

the down-regulation in clear presence vs. absence situation as in the *in vitro* negative control was about 20-fold (see Figure 4.41). The absolute numbers obtained by MS_QBAT quantification might even be underestimated due to quantification against an artificial background values in a presence vs. absence situation. The increased background fraction in the kidney samples diminishes the median ratio values. Within the liver samples, no clear trends in down-regulation are visible, as the background fraction is heavily increased (further elucidated in the previous chapter). Also the closer analysis of proteins annotated to the plasma membrane and the extracellular matrix confirmed the decreased reactivity, as more than 60% of the proteins were more than 2-fold up-regulated with commercial Sulfo-NHS-LC-biotin (compared to 74% in the PBS-perfused and 85% in the non-perfused negative control). The same trend was also visible in the SRM-based quantification of plasma membrane annotated protein candidates (see Figures 4.46, 4.54). The fraction of uniquely identified plasma membrane or extracellular matrix proteins in comparison to commercial Sulfo-NHS-LC-biotin is low with 4%/8% with Biotin-Heparin-sNHS and Biotin-Enoxaparin-sNHS, respectively, against 29%/19% with Sulfo-NHS-LC-biotin, respectively (analysis without one hit wonders). Hereby, Biotin-Enoxaparin-sNHS might be slightly more reactive compared to Biotin-Heparin-sNHS due to the smaller size of the reagent. Besides, cluster-analysis of the quantified proteins has shown the similarity of the samples biotinylated with Biotin-Heparin-sNHS or Biotin-Enoxaparin-sNHS to the negative controls. Decreased reactivity leads to an artificial overestimation of the identified intracellular background proteome fraction, e.g. in the analysis of subcellular localization fractions, and to the high similarity to the negative controls e.g. in cluster analysis.

Furthermore, with the peptide-based reagents, it was not possible to achieve the full reactivity of commercial Sulfo-NHS-LC-biotin, which was shown earlier by reactivity assessment on BSA and on HeLa cells. The effect is less pronounced in mass spectrometric analysis of tissue samples, on the one hand due to the application of a more favorable excess of the biotinylation reagents (25 μ mol per mouse) and on the other hand due to artificial variability reduction by the increased background proteome fraction. The changes of the linker structure by increase in size and introduction of charged groups to improve the reagents' selectivity is at the same time responsible for a reactivity decrease. Bulky side groups as well as reagents with huge linkers are sterically more hindered during the labelling reaction. This is why, in the mass spectrometric datasets, SMCC-Cys-(L-Asp)₃-biotin is slightly more reactive than the more spacious SM(PEG)₆-Cys-(L-Asp)₃-biotin. The same is true for Biotin-Heparin-sNHS and Biotin-Enoxaparin-sNHS, and both large-scaled heparin-based reagents exhibit a heavily decreased reactivity compared to all other biotinylation reagents.

The tissue datasets have shown that both peptide-based and commercial reagents identify comparable fractions in terms of total protein numbers and signal intensities of the proteotypic peptides. Analysis of the identified plasma membrane or extracellular matrix annotated fraction in the kidney dataset results in comparable fractions uniquely identified with the different biotinylation reagents (6% to 8%,

without one hit wonders). Also the quantification against commercial Sulfo-NHS-LC-biotin proves no differences in terms of stability of target proteome identification. About 40% of the quantified proteins are more than 2-fold up- or down-regulated, but (unlike the heparin-based reagents) in comparable fractions with one or the other reagent (see Figure 4.44). High numbers of identified peptides per protein increases the stability of the quantification. The number of identified proteotypic peptides per plasma membrane or extracellular matrix annotated protein is significantly increased with both peptide-based and commercial reagents compared to intracellular proteins (see Figure 4.36). Highest numbers were hereby obtained with the peptide-based reagents. Within the liver dataset, differences across the groups are small, also at the level of the number of identified proteotypic peptides per protein. Only the number of proteotypic peptides identified for the plasma membrane/extracellular fraction is slightly, but non-significantly reduced within the negative controls. Reasons for the artificial variability decrease by increased background proteome identification have already been discussed. Despite the co-identified background proteome fraction, the enrichment of vascular accessible surface and extracellular matrix proteins works reliably and successful with the peptide-based reagents, as additionally shown by SRM analysis of kidney and liver samples. Most low- or medium-abundant plasma membrane target proteins could not be quantified out of full proteome samples, whereas preceding enrichment with peptide-based and commercial reagents provided a technology platform to enable their stable quantification.

Cluster analysis of all reagent and control groups first of all grouped samples according to the reagent's reactivity (based on the similarly identified background proteome fractions), and secondly based on the reagent's linker type (see Figures 4.42, 4.52). The interplay between selectivity and reactivity presents the main challenge in the work with novel biotinylation reagents. A more selective biotinylation reagent might not identify more plasma membrane proteins (or less intracellular proteins) than a less selective one, if it exhibits a decreased reactivity. Also the dataset published by Strassberger *et al.* suffers from the low reactivity of the NHS- β -Ala-(L-Asp)₃-biotin: Only 364 and 392 proteins in total could be identified in the kidney and liver dataset, hereby the reported fraction of proteins with plasma membrane or extracellular matrix annotation was 60% (219 proteins) and 42% (166 proteins) respectively.^[12] Within this thesis, the double amount of membrane or extracellular matrix annotated proteins could be stably identified and quantified (without one hit wonders) with the reactivity-improved peptide-based reagents.

To summarize the findings from the validation work with the novel biotinylation reagents, it can be stated that an ideal reagent for biomarker discovery studies would be site-specifically activated, not too big, with non-bulky side-groups and low steric hindrance for high reactivity, but highly charged and not too small to prevent plasma membrane crossing. In reality, every novel biotinylation reagent is a compromise of those properties. Furthermore, it has been shown that optimized sample preparation strategies are the basis for successful plasma membrane proteome enrichment. The fraction of necrotic or apoptotic cells has to be minimized and delipidation is a crucial step for the work with tissue samples.

6. Conclusions and Outlook

Within this thesis, two novel peptide-based biotinylation reagents (SMCC-Cys-(L-Asp)₃-biotin and SM(PEG)₆-Cys-(L-Asp)₃-biotin) were developed, which significantly improved the reactivity of non-specifically activated NHS- β -Ala-(L-Asp)₃-biotin published by Strassberger *et al.*^[12] Additionally, two novel biotinylation reagents based on a heparin linker (Biotin-Heparin-sNHS and Biotin-Enoxaparin-sNHS) large in size and multiply charged were synthesized and validated. Furthermore, two alkyne-based reagents for bioorthogonal click-chemistry mediated capturing, with comparable reactivity to the novel peptide-based reagents, have been synthesised.

The mass spectrometric analyses performed within this thesis emphasize the importance of an optimized sample preparation workflow for successful surface proteome enrichment as well as the crucial interplay between selectivity and reactivity for the design of novel biotinylation reagents. Heparin-based reagents exhibited a significantly reduced reactivity, whereas the peptide-based reagents provide a stable technology platform for the identification and quantification of plasma membrane or extracellular matrix proteins in mass spectrometric experiments. Even low-abundant proteins, which cannot be identified out of full proteome lysate samples, are stably identified and quantified with many proteotypic peptides in discovery experiments or can be targeted via SRM-based quantification. Maintenance of maximal cell viability as well as delipidation of tissue samples turned out to be crucial for successful identification of a large fraction of plasma membrane or extracellular matrix annotated proteins.

One of the challenges when working with plasma membrane proteins is the lack of polar tryptic cleavages sites in the transmembrane domains.^{[290],[119]} Sequence coverage of the plasma membrane proteins could be further improved by combination of tryptic digestion with another digestion enzyme or a chemical cleavage method. Tryptic digest followed by cyanogen bromide cleavage (C-terminal of methionine residues) has been shown to double the cleavage sites in hydrophobic transmembrane domains.^[142]

To further assess the differences in labelling with the peptide-based and commercial biotinylation reagents, the application of higher reagent amounts have to be evaluated. However, the applied amount must not interfere with cell viability. Reactivity-dependent effects would be decreased, and therefore also the influence of the identified background proteome fraction on the data analysis. To enable direct comparison with commercial reagents, both in terms of reactivity and in the fraction of identified plasma membrane proteins, the effects should be assessed with varying reagent amounts of all biotinylation reagents.

To identify differences in the chemical surroundings of the targeted primary amino groups, it is necessary to analyse the labelled peptide fragment that remains bound to the streptavidin-sepharose after tryptic digest. Differences in total protein size or polarity are not meaningful enough as only a small protein part is accessible for the biotinylation reagents on the cell surface. Furthermore, sequence coverage of the labelled proteins is increased, if the labelled peptide fragments can be included in the analysis. Two approaches are possible to release the biotinylated peptide fragment from the streptavidin-sepharose. The first possibility is the disruption of the non-covalent biotin-streptavidin interaction under harsh elution conditions by competition with free Biotin.^[418] The second possibility is the re-design of the biotinylation reagents: Introduction of a cleavable linker (e.g. a vicinal diol group cleavable by mild oxidative treatment) would provide the possibility of a chemical release of the peptide fragments.^[191]

The identification of robust, vascular-accessible biomarkers is a prerequisite for the development of novel therapeutic or diagnostic tools. Vascular accessible targets can be located on the surface of diseased cells, in the surrounding extracellular matrix or on the surface of newly formed blood vessels: Angiogenic events in a healthy adult are rare, but reported under various pathological conditions such as tumour formation, inflammatory diseases, atherosclerosis or upon transplantations.^{[14],[94]} The targeted delivery of bioactive molecules directly to the site of disease reduces side-effects in healthy tissues. Diagnostic or therapeutic antibodies or antibody fragments can hereby either interfere with a target on the diseased cells that is essential for growth or survival or target a biomarker on the cell or in the cell surroundings while transporting a therapeutic payload such as cytotoxic drugs, radionuclides or cytokines.^[419] The novel peptide-based reagents provide a reliable technology platform for the enrichment, identification and quantification of surface or vascular accessible proteins. Hereby, biotinylation can be performed *in vitro* by biotinylation of cells, via *in vivo* perfusion of rodents or *ex vivo* with surgically resected tissue material. The success of these biomarker studies is depending on the vascularisation of the target tissue as well as on an optimized sample preparation protocol. Several delipidation steps might improve the reduction of the intracellular background proteome identified in mass spectrometric analyses of liver tissue. Novel plasma membrane or extracellular matrix annotated targets can be identified by stable quantification of diseased vs. healthy tissue based on many stably identified proteotypic peptides. Furthermore, successful SRM-based quantification of potential medium- to low-abundant targets is enabled by previous surface proteome enrichment.

List of Figures

1.1. Biomarker discovery on different molecular levels.	29
1.2. Different classes of membrane proteins.	36
1.3. Technologies for cell surface proteome enrichment.	42
1.4. Typical shotgun proteomics workflow.	43
1.5. Mass spectrometric shotgun analysis.	45
1.6. Nomenclature of peptide fragment ions.	46
1.7. Main principles of label-free quantification.	49
1.8. Mass spectrometric SWATH analysis.	51
1.9. Mass spectrometric SRM analysis.	53
4.1. Commercial biotinylation reagents.	93
4.2. NHS- β -Ala-(L-Asp) ₃ -biotin.	94
4.3. Biotinylation of BSA with Sulfo-NHS-LC-biotin under different conditions	95
4.4. Reactivity comparison of NHS- β -Ala-(L-Asp) ₃ -biotin and Sulfo-NHS-LC-biotin.	96
4.5. Site-unspecific activation of Biotin-(L-Asp) ₃ - β -Ala.	97
4.6. Novel peptide-based biotinylation reagents.	98
4.7. Reactivity comparison of novel and commercial reagents.	99
4.8. Biotinylation with SMCC-Cys-(L-Asp) ₃ -biotin under different conditions.	99
4.9. Structure of heparin.	100
4.10. Biotinylation of BSA with Enoxaparin-sNHS.	102
4.11. Novel click-chemistry-based reagents.	104
4.12. Reactivity assesement of SMCC-Cys-(L-Asp) ₃ -PA-alkyne and SMCC-Cys-(L-Asp) ₃ -DBCO-alkyne.	105
4.13. Workflow <i>in vitro</i> validation.	106
4.14. FACS analysis of Hela biotinylated with NHS- β -Ala-(L-Asp) ₃ -biotin in comparison to Sulfo-NHS-LC-biotin and SMCC-Cys-(L-Asp) ₃ -biotin.	107
4.15. FACS analysis of Hela biotinylated with novel peptide-based reagents.	107
4.16. CLSM of biotinylated HeLa cells.	108
4.17. FACS analysis of Hela biotinylated with novel heparin-based reagents.	108
4.18. FACS analysis of Hela biotinylated for mass spectrometric analysis. .	109
4.19. SDS-PAGE and Western Blot of biotinylated cell lysate and supernatant after capturing.	110
4.20. 2D peptide maps of <i>in vitro</i> biotinylated samples.	111
4.21. Venn diagram of identified proteins, <i>in vitro</i> dataset.	112

4.22. Enrichment analysis (cellular component) of proteins identified with all biotinylation reagents, <i>in vitro</i> dataset.	115
4.23. Protein localization in the <i>in vitro</i> dataset.	116
4.24. Annotation propagation with MS _Q BAT, <i>in vitro</i> dataset.	118
4.25. Quantification of pairs of technical replicates, <i>in vitro</i> dataset.	118
4.26. Quantification of all biotinylation reagents vs. Sulfo-NHS-LC-biotin, <i>in vitro</i> dataset.	119
4.27. Protein localization in the heparin <i>in vitro</i> dataset.	120
4.28. Enrichment analysis (cellular component, molecular function) of proteins identified with heparin-based reagents, <i>in vitro</i> dataset.	121
4.29. Workflow <i>in vivo</i> validation.	124
4.30. Visualisation of biotinylation depth.	126
4.31. CLSM of mouse tissue perfused with commercial and peptide-based reagents.	127
4.32. CLSM of mouse tissue perfused with heparin-based reagents.	128
4.33. ELISA of biotinylated mouse kidney and liver tissue.	129
4.34. Western Blots of biotinylated kidney homogenates.	130
4.35. Protein localization in the kidney dataset.	132
4.36. Average number of proteotypic peptides per identified protein in the kidney dataset.	134
4.37. Volcano plots of technical replicates, kidney dataset.	135
4.38. Volcano plots of biological replicates, kidney dataset.	135
4.39. Volcano plots of super-samples from all reagents against Sulfo-NHS-LC-biotin, kidney dataset.	136
4.40. Quantification of technical and biological replicates or between reagent groups, kidney dataset.	136
4.41. Quantification of all biotinylation reagents vs. Sulfo-NHS-LC-biotin, kidney dataset.	137
4.42. Cluster analysis kidney dataset.	138
4.43. Analysis of the quantified plasma membrane/extracellular fraction, kidney dataset.	139
4.44. Regulation analysis of the quantified plasma membrane/extracellular fraction, kidney dataset.	139
4.45. Enrichment analysis (cellular component) of the top-300 proteins identified in PBS-perfused mouse kidney.	143
4.46. SRM analysis of kidney samples.	146
4.47. Protein localization in the liver dataset.	148
4.48. Average number of proteotypic peptides per identified protein in the liver dataset.	149
4.49. Volcano plots of biological replicates, liver dataset.	149
4.50. Volcano plots of super-samples from all reagents against Sulfo-NHS-LC-biotin, liver dataset.	150
4.51. Quantification of all biotinylation reagents vs. Sulfo-NHS-LC-biotin, liver dataset.	150

4.52. Cluster analysis liver dataset.	151
4.53. Enrichment analysis (cellular component) of 605 proteins identified with all reagents and the negative controls, liver dataset.	152
4.54. SRM analysis of liver samples.	155
A.1. CLSM staining controls of mouse tissue perfused with commercial, peptide- and heparin-based reagents.	220
A.2. CLSM of mouse kidney samples perfused with commercial and peptide- based reagents used for mass spectrometric <i>in vivo</i> validation.	221
A.3. CLSM of mouse kidney samples perfused with heparin-based reagents used for mass spectrometric <i>in vivo</i> validation.	222
A.4. CLSM staining controls of kidney tissue perfused with commercial, peptide- and heparin-based reagents.	223
A.5. CLSM of mouse liver samples perfused with commercial and peptide- based reagents used for mass spectrometric <i>in vivo</i> validation.	224
A.6. CLSM of mouse liver samples perfused with heparin-based reagents used for mass spectrometric <i>in vivo</i> validation.	225
A.7. CLSM staining controls of liver tissue perfused with commercial, peptide- and heparin-based reagents.	226
A.8. 2D peptide maps of <i>in vivo</i> biotinylated kidney samples.	227
A.9. 2D peptide maps of <i>in vivo</i> biotinylated liver samples.	228
A.10. Background-reduced peak areas of the heavy standards obtained by SRM analysis.	229

List of Tables

3.1. RIPA buffer for cell lysis.	72
3.2. Protein extraction buffer for tissue samples.	75
3.3. 5x Gel loading buffer.	77
3.4. 20x Western blot transfer buffer.	77
3.5. UPLC gradient used for discovery experiment.	81
3.6. Internal standard peptides.	82
3.7. Standard peptides for calibration.	83
3.8. UPLC gradient used for SRM experiment.	85
4.1. Top hit proteins identified in the negative control, <i>in vitro</i> dataset. . .	112
4.2. Subunits of the 20S core of the 26S proteasome identified in the <i>in vitro</i> dataset.	113
4.3. Plasma membrane proteins (examples) identified in the <i>in vitro</i> dataset.	114
4.4. Sample preparation tests on cells biotinylated with Biotin-Heparin-sNHS.	122
4.5. Overlap of the plasma membrane/extracellular protein fraction between different sample groups vs. Sulfo-NHS-LC-biotin, kidney dataset.	133
4.6. Sample preparation tests on kidney negative control samples.	141
B.1. Transition list SRM analysis, kidney tissue.	231
B.2. Transition list SRM analysis, liver tissue.	231
B.3. <i>In vitro</i> validation of peptide-based reagents, analysis of peptide summaries.	231
B.4. <i>In vitro</i> validation of peptide-based reagents, relative quantification. .	231
B.5. <i>In vitro</i> validation of heparin-based reagents, analysis of peptide summaries.	231
B.6. <i>In vivo</i> validation, kidney tissue, analysis of peptide summaries. . . .	231
B.7. <i>In vivo</i> validation, kidney tissue, relative quantification.	231
B.8. <i>In vivo</i> background proteome, analysis of peptide summaries.	232
B.9. <i>In vivo</i> validation, liver tissue, analysis of peptide summaries.	232
B.10. <i>In vivo</i> validation, liver tissue, relative quantification.	232

List of Schemes

4.1. Protein modification with NHS-ester derivatives.	92
4.2. Activation of novel peptide-based reagents.	97
4.3. Biotinylation of heparin.	101
4.4. Carboxylic group activation at heparin side chains.	101
4.5. Protein modification and click-chemistry-mediated capturing.	103

Bibliography

- [1] Biomarker Definitions Working Group. “Biomarkers and surrogate endpoints: preferred definitions and conceptual framework”. In: *Clin. Pharmacol. Ther.* 69.3 (Mar. 2001), pp. 89–95.
- [2] A. de Gramont, S. Watson, L. M. Ellis, J. Rodón, J. Tabernero, A. de Gramont, and S. R. Hamilton. “Pragmatic issues in biomarker evaluation for targeted therapies in cancer”. In: *Nat. Rev. Clin. Oncol.* 12.4 (Apr. 2015), pp. 197–212.
- [3] *Environmental Health Criteria 222: Biomarkers in Risk Assessment: Validity and Validation*. WHO. 2001.
- [4] K. Strimbu and J. A. Tavel. “What are biomarkers?” In: *Curr. Opin. HIV AIDS* 5.6 (Nov. 2010), pp. 463–466.
- [5] J. J. Smith, A. G. Sorensen, and J. H. Thrall. “Biomarkers in imaging: realizing radiology’s future”. In: *Radiology* 227.3 (June 2003), pp. 633–638.
- [6] D. F. Hayes, R. C. Bast, C. E. Desch, H. Fritsche, N. E. Kemeny, J. M. Jessup, G. Y. Locker, J. S. Macdonald, R. G. Mennel, L. Norton, P. Ravdin, S. Taube, and R. J. Winn. “Tumor marker utility grading system: a framework to evaluate clinical utility of tumor markers”. In: *J. Natl. Cancer Inst.* 88.20 (Oct. 1996), pp. 1456–1466.
- [7] S. Mehta, A. Shelling, A. Muthukaruppan, A. Lasham, C. Blenkiron, G. Laking, and C. Print. “Predictive and prognostic molecular markers for cancer medicine”. In: *Ther. Adv. Med. Oncol.* 2.2 (Mar. 2010), pp. 125–148.
- [8] J. Adamski and K. Suhre. “Metabolomics platforms for genome wide association studies—linking the genome to the metabolome”. In: *Curr. Opin. Biotechnol.* 24.1 (Feb. 2013), pp. 39–47.
- [9] E. Drucker and K. Krapfenbauer. “Pitfalls and limitations in translation from biomarker discovery to clinical utility in predictive and personalised medicine”. In: *EPMA J.* 4.1 (2013), p. 7.
- [10] V. Brower. “Biomarkers: Portents of malignancy”. In: *Nature* 471.7339 (Mar. 2011), pp. 19–21.
- [11] “Targeted Therapy”. In: *Dictionary of Cancer Terms*. May 2015. URL: <http://www.cancer.gov/publications/dictionaries/cancer-terms?cdrid=270742> (visited on 07/01/2015).

- [12] V. Strassberger, S. Trüssel, T. Fugmann, D. Neri, and C. Rösli. “A novel reactive ester derivative of biotin with reduced membrane permeability for in vivo biotinylation experiments”. In: *Proteomics* 10.19 (Oct. 2010), pp. 3544–3548.
- [13] C. Rösli and D. Neri. “Methods for the identification of vascular markers in health and disease: from the bench to the clinic”. In: *J. Proteomics* 73.11 (Oct. 2010), pp. 2219–2229.
- [14] F. Bootz, A. S. Schmid, and D. Neri. “Alternatively Spliced EDA Domain of Fibronectin Is a Target for Pharmacodelivery Applications in Inflammatory Bowel Disease”. In: *Inflamm. Bowel Dis.* (May 2015).
- [15] F. Doll, K. Schwager, T. Hemmerle, and D. Neri. “Murine analogues of etanercept and of F8-IL10 inhibit the progression of collagen-induced arthritis in the mouse”. In: *Arthritis Res. Ther.* 15.5 (2013), R138.
- [16] U. H. Weidle, D. Maisel, S. Klostermann, C. Schiller, and E. H. Weiss. “Intracellular proteins displayed on the surface of tumor cells as targets for therapeutic intervention with antibody-related agents”. In: *Cancer Genomics Proteomics* 8.2 (2011), pp. 49–63.
- [17] M. H. Hansen, H. V. Nielsen, and H. J. Ditzel. “Translocation of an intracellular antigen to the surface of medullary breast cancer cells early in apoptosis allows for an antigen-driven antibody response elicited by tumor-infiltrating B cells”. In: *J. Immunol.* 169.5 (Sept. 2002), pp. 2701–2711.
- [18] Q. Yu and I. Stamenkovic. “Cell surface-localized matrix metalloproteinase-9 proteolytically activates TGF-beta and promotes tumor invasion and angiogenesis”. In: *Genes Dev.* 14.2 (Jan. 2000), pp. 163–176.
- [19] M. Ferrarini, S. Heltai, M. R. Zocchi, and C. Rugarli. “Unusual expression and localization of heat-shock proteins in human tumor cells”. In: *Int. J. Cancer* 51.4 (June 1992), pp. 613–619.
- [20] Y. Ding, N. Song, C. Liu, T. He, W. Zhuo, X. He, Y. Chen, X. Song, Y. Fu, and Y. Luo. “Heat shock cognate 70 regulates the translocation and angiogenic function of nucleolin”. In: *Arterioscler. Thromb. Vasc. Biol.* 32.9 (Sept. 2012), e126–134.
- [21] D. Nedelkov, U. A. Kiernan, E. E. Niederkofler, K. A. Tubbs, and R. W. Nelson. “Investigating diversity in human plasma proteins”. In: *Proc. Natl. Acad. Sci. U.S.A.* 102.31 (Aug. 2005), pp. 10852–10857.
- [22] M. I. Maqsood, M. M. Matin, A. R. Bahrami, and M. M. Ghasroldasht. “Immortality of cell lines: challenges and advantages of establishment”. In: *Cell Biol. Int.* 37.10 (Oct. 2013), pp. 1038–1045.
- [23] D. M. Peehl. “Primary cell cultures as models of prostate cancer development”. In: *Endocr. Relat. Cancer* 12.1 (Mar. 2005), pp. 19–47.

- [24] D.-J. Cheon and S. Orsulic. “Mouse models of cancer”. In: *Annu. Rev. Pathol.* 6 (2011), pp. 95–119.
- [25] M. P. Kim, D. B. Evans, H. Wang, J. L. Abbruzzese, J. B. Fleming, and G. E. Gallick. “Generation of orthotopic and heterotopic human pancreatic cancer xenografts in immunodeficient mice”. In: *Nat. Protoc.* 4.11 (2009), pp. 1670–1680.
- [26] A. Richmond and Y. Su. “Mouse xenograft models vs GEM models for human cancer therapeutics”. In: *Dis. Model Mech.* 1.2-3 (2008), pp. 78–82.
- [27] E. R. LaVallie, A. J. Dorner, and M. E. Burczynski. “Use of ex vivo systems for biomarker discovery”. In: *Curr. Opin. Pharmacol.* 8.5 (Oct. 2008), pp. 647–653.
- [28] The Cancer Genome Atlas Research Network. “Comprehensive genomic characterization defines human glioblastoma genes and core pathways”. In: *Nature* 455.7216 (Oct. 2008), pp. 1061–1068.
- [29] M. Liu, L. Jiang, and X.-Y. Guan. “The genetic and epigenetic alterations in human hepatocellular carcinoma: a recent update”. In: *Protein Cell* 5.9 (Sept. 2014), pp. 673–691.
- [30] F. Sanger and A. R. Coulson. “A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase”. In: *J. Mol. Biol.* 94.3 (May 1975), pp. 441–448.
- [31] F. Sanger, S. Nicklen, and A. R. Coulson. “DNA sequencing with chain-terminating inhibitors”. In: *Proc. Natl. Acad. Sci. U.S.A.* 74.12 (Dec. 1977), pp. 5463–5467.
- [32] A. von Bubnoff. “Next-generation sequencing: the race is on”. In: *Cell* 132.5 (Mar. 2008), pp. 721–723.
- [33] M. Margulies et al. “Genome sequencing in microfabricated high-density picolitre reactors”. In: *Nature* 437.7057 (Sept. 2005), pp. 376–380.
- [34] R. L. Momparker. “Cancer epigenetics”. In: *Oncogene* 22.42 (Sept. 2003), pp. 6479–6483.
- [35] K. Mensaert, S. Denil, G. Trooskens, W. Van Criekinge, O. Thas, and T. De Meyer. “Next-generation technologies and data analytical approaches for epigenomics”. In: *Environ. Mol. Mutagen.* 55.3 (Apr. 2014), pp. 155–170.
- [36] Z. Wang, M. Gerstein, and M. Snyder. “RNA-Seq: a revolutionary tool for transcriptomics”. In: *Nat. Rev. Genet.* 10.1 (Jan. 2009), pp. 57–63.
- [37] J. DeRisi, L. Penland, P. O. Brown, M. L. Bittner, P. S. Meltzer, M. Ray, Y. Chen, Y. A. Su, and J. M. Trent. “Use of a cDNA microarray to analyse gene expression patterns in human cancer”. In: *Nat. Genet.* 14.4 (Dec. 1996), pp. 457–460.

- [38] V. E. Velculescu, L. Zhang, B. Vogelstein, and K. W. Kinzler. “Serial analysis of gene expression”. In: *Science* 270.5235 (Oct. 1995), pp. 484–487.
- [39] K. L. Nielsen, A. L. Høgh, and J. Emmersen. “DeepSAGE–digital transcriptomics with high sensitivity, simple experimental protocol and multiplexing of samples”. In: *Nucleic Acids Res.* 34.19 (2006), e133.
- [40] Y. Zhang, P. Yang, and X. F. Wang. “Microenvironmental regulation of cancer metastasis by miRNAs”. In: *Trends Cell Biol.* 24.3 (Mar. 2014), pp. 153–160.
- [41] R. A. Zubarev. “The challenge of the proteome dynamic range and its implications for in-depth proteomics”. In: *Proteomics* 13.5 (Mar. 2013), pp. 723–726.
- [42] B. Schwanhäusser, D. Busse, N. Li, G. Dittmar, J. Schuchhardt, J. Wolf, W. Chen, and M. Selbach. “Global quantification of mammalian gene expression control”. In: *Nature* 473.7347 (May 2011), pp. 337–342.
- [43] M. Wilhelm et al. “Mass-spectrometry-based draft of the human proteome”. In: *Nature* 509.7502 (May 2014), pp. 582–587.
- [44] S. Hanash. “Disease proteomics”. In: *Nature* 422.6928 (Mar. 2003), pp. 226–232.
- [45] R. Wilson. “Sensitivity and specificity: twin goals of proteomics assays. Can they be combined?” In: *Expert Rev. Proteomics* 10.2 (Apr. 2013), pp. 135–149.
- [46] O. Poetz, J. M. Schwenk, S. Kramer, D. Stoll, M. F. Templin, and T. O. Joos. “Protein microarrays: catching the proteome”. In: *Mech. Ageing Dev.* 126.1 (Jan. 2005), pp. 161–170.
- [47] P. Tighe, O. Negm, I. Todd, and L. Fairclough. “Utility, reliability and reproducibility of immunoassay multiplex kits”. In: *Methods* 61.1 (May 2013), pp. 23–29.
- [48] X. Han and R. W. Gross. “Shotgun lipidomics: electrospray ionization mass spectrometric analysis and quantitation of cellular lipidomes directly from crude extracts of biological samples”. In: *Mass. Spectrom. Rev.* 24.3 (2005), pp. 367–412.
- [49] B. Brugger. “Lipidomics: analysis of the lipid composition of cells and sub-cellular organelles by electrospray ionization mass spectrometry”. In: *Annu. Rev. Biochem.* 83 (2014), pp. 79–98.
- [50] M. R. Wenk. “The emerging field of lipidomics”. In: *Nat. Rev. Drug Discov.* 4.7 (July 2005), pp. 594–610.
- [51] M. R. Wenk. “Lipidomics: new tools and applications”. In: *Cell* 143.6 (Dec. 2010), pp. 888–895.

- [52] C. R. Santos and A. Schulze. “Lipid metabolism in cancer”. In: *FEBS J.* 279.15 (Aug. 2012), pp. 2610–2623.
- [53] X. Zhou, J. Mao, J. Ai, Y. Deng, M. R. Roth, C. Pound, J. Henegar, R. Welti, and S. A. Bigler. “Identification of plasma lipid biomarkers for prostate cancer by lipidomics and bioinformatics”. In: *PLoS ONE* 7.11 (2012), e48889.
- [54] C. Zhao, J. Mao, J. Ai, M. Shenwu, T. Shi, D. Zhang, X. Wang, Y. Wang, and Y. Deng. “Integrated lipidomics and transcriptomic analysis of peripheral blood reveals significantly enriched pathways in type 2 diabetes mellitus”. In: *BMC Med. Genomics* 6 Suppl 1 (2013), S12.
- [55] C. L. Kien, J. Y. Bunn, M. E. Poynter, R. Stevens, J. Bain, O. Ikayeva, N. K. Fukagawa, C. M. Champagne, K. I. Crain, T. R. Koves, and D. M. Muoio. “A lipidomics analysis of the relationship between dietary fatty acid composition and insulin sensitivity in young adults”. In: *Diabetes* 62.4 (Apr. 2013), pp. 1054–1063.
- [56] P. L. Wood. “Lipidomics of Alzheimer’s disease: current status”. In: *Alzheimers Res. Ther.* 4.1 (2012), p. 5.
- [57] T. Seppänen-Laakso and M. Orešič. “How to study lipidomes”. In: *J. Mol. Endocrinol.* 42.3 (Mar. 2009), pp. 185–190.
- [58] M. J. Gerl, T. Sachsenheimer, M. Grzybek, U. Coskun, F. T. Wieland, and B. Brugger. “Analysis of transmembrane domains and lipid modified peptides with matrix-assisted laser desorption ionization-time-of-flight mass spectrometry”. In: *Anal. Chem.* 86.8 (Apr. 2014), pp. 3722–3726.
- [59] M. R. Wenk. “Lipidomics in drug and biomarker development”. In: *Expert Opin. Drug Discov.* 1.7 (Dec. 2006), pp. 723–736.
- [60] D. Gode and D. A. Volmer. “Lipid imaging by mass spectrometry - a review”. In: *Analyst* 138.5 (Mar. 2013), pp. 1289–1315.
- [61] R. L. Schnaar. “Glycans and glycan-binding proteins in immune regulation: A concise introduction to glycobiology for the allergist”. In: *J. Allergy Clin. Immunol.* 135.3 (Mar. 2015), pp. 609–615.
- [62] R. Apweiler, H. Hermjakob, and N. Sharon. “On the frequency of protein glycosylation, as deduced from analysis of the SWISS-PROT database”. In: *Biochim. Biophys. Acta* 1473.1 (Dec. 1999), pp. 4–8.
- [63] R. Raman, S. Raguram, G. Venkataraman, J. C. Paulson, and R. Sasisekharan. “Glycomics: an integrated systems approach to structure-function relationships of glycans”. In: *Nat. Methods* 2.11 (Nov. 2005), pp. 817–824.
- [64] B. Adamczyk, T. Tharmalingam, and P. M. Rudd. “Glycans as cancer biomarkers”. In: *Biochim. Biophys. Acta* 1820.9 (Sept. 2012), pp. 1347–1353.

- [65] S. A. Brooks, T. M. Carter, L. Royle, D. J. Harvey, S. A. Fry, C. Kinch, R. A. Dwek, and P. M. Rudd. “Altered glycosylation of proteins in cancer: what is the potential for new anti-tumour strategies”. In: *Anticancer Agents Med. Chem.* 8.1 (Jan. 2008), pp. 2–21.
- [66] J. Zaia. “Glycosaminoglycan glycomics using mass spectrometry”. In: *Mol. Cell. Proteomics* 12.4 (Apr. 2013), pp. 885–892.
- [67] T. Feizi, F. Fazio, W. Chai, and C. H. Wong. “Carbohydrate microarrays - a new set of technologies at the frontiers of glycomics”. In: *Curr. Opin. Struct. Biol.* 13.5 (Oct. 2003), pp. 637–645.
- [68] J. C. Lindon, E. Holmes, and J. K. Nicholson. “Metabonomics in pharmaceutical R and D”. In: *FEBS J.* 274.5 (Mar. 2007), pp. 1140–1151.
- [69] J. J. Ramsden. “Bioinformatics”. In: Springer-Verlag, 2009. Chap. 16. Metabolomics and Metabonomics.
- [70] V. Shulaev. “Metabolomics technology and bioinformatics”. In: *Brief. Bioinformatics* 7.2 (June 2006), pp. 128–139.
- [71] R. Goodacre. “Metabolomics of a superorganism”. In: *J. Nutr.* 137.1 Suppl (Jan. 2007), 259S–266S.
- [72] M. R. Smith. “Rituximab (monoclonal anti-CD20 antibody): mechanisms of action and resistance”. In: *Oncogene* 22.47 (Oct. 2003), pp. 7359–7368.
- [73] P. Chames, M. Van Regenmortel, E. Weiss, and D. Baty. “Therapeutic antibodies: successes, limitations and hopes for the future”. In: *Br. J. Pharmacol.* 157.2 (May 2009), pp. 220–233.
- [74] G. M. Thurber, M. M. Schmidt, and K. D. Wittrup. “Factors determining antibody distribution in tumors”. In: *Trends Pharmacol. Sci.* 29.2 (Feb. 2008), pp. 57–61.
- [75] R. K. Jain and L. T. Baxter. “Mechanisms of heterogeneous distribution of monoclonal antibodies and other macromolecules in tumors: significance of elevated interstitial pressure”. In: *Cancer Res.* 48.24 Pt 1 (Dec. 1988), pp. 7022–7032.
- [76] T. P. Padera, B. R. Stoll, J. B. Tooredman, D. Capen, E. di Tomaso, and R. K. Jain. “Pathology: cancer cells compress intratumour vessels”. In: *Nature* 427.6976 (Feb. 2004), p. 695.
- [77] J. Denekamp and B. Hobson. “Endothelial-cell proliferation in experimental tumours”. In: *Br. J. Cancer* 46.5 (Nov. 1982), pp. 711–720.
- [78] A. I. Minchinton and I. F. Tannock. “Drug penetration in solid tumours”. In: *Nat. Rev. Cancer* 6.8 (Aug. 2006), pp. 583–592.
- [79] R. K. Jain. “Normalization of tumor vasculature: an emerging concept in antiangiogenic therapy”. In: *Science* 307.5706 (Jan. 2005), pp. 58–62.

- [80] D. Bouÿs, G. A. P. Hospers, C. Meijer, G. Molema, and N. H. Mulder. “Endothelium in vitro: a review of human vascular endothelial cell lines for blood vessel-related research”. In: *Angiogenesis* 4.2 (2001), pp. 91–102.
- [81] C. Michiels. “Endothelial cell functions”. In: *J. Cell. Physiol.* 196.3 (Sept. 2003), pp. 430–443.
- [82] W. Risau. “Mechanisms of angiogenesis”. In: *Nature* 386.6626 (Apr. 1997), pp. 671–674.
- [83] D. Hanahan and J. Folkman. “Patterns and emerging mechanisms of the angiogenic switch during tumorigenesis”. In: *Cell* 86.3 (Aug. 1996), pp. 353–364.
- [84] S. M. Weis and D. A. Cheresh. “Tumor angiogenesis: molecular pathways and therapeutic targets”. In: *Nat. Med.* 17.11 (2011), pp. 1359–1370.
- [85] T. Udagawa, A. Fernandez, E.-G. Achilles, J. Folkman, and R. J. D’Amato. “Persistence of microscopic human cancers in mice: alterations in the angiogenic balance accompanies loss of tumor dormancy”. In: *FASEB J.* 16.11 (Sept. 2002), pp. 1361–1370.
- [86] J. Folkman. “Angiogenesis”. In: *Annu. Rev. Med.* 57 (2006), pp. 1–18.
- [87] G. Bergers and L. E. Benjamin. “Tumorigenesis and the angiogenic switch”. In: *Nat. Rev. Cancer* 3.6 (June 2003), pp. 401–410.
- [88] A. C. Dudley. “Tumor endothelial cells”. In: *Cold Spring Harb. Perspect. Med.* 2.3 (Mar. 2012), a006536.
- [89] B. St. Croix, C. Rago, V. Velculescu, G. Traverso, K. E. Romans, E. Montgomery, A. Lal, G. J. Riggins, C. Lengauer, B. Vogelstein, and K. W. Kinzler. “Genes expressed in human tumor endothelium”. In: *Science* 289.5482 (Aug. 2000), pp. 1197–1202.
- [90] P. Oh, Y. Li, J. Yu, E. Durr, K. M. Krasinska, L. A. Carver, J. E. Testa, and J. E. Schnitzer. “Subtractive proteomic mapping of the endothelial surface in lung and solid tumours for tissue-specific therapy”. In: *Nature* 429.6992 (June 2004), pp. 629–635.
- [91] D. Neri and R. Bicknell. “Tumour vascular targeting”. In: *Nat. Rev. Cancer* 5.6 (June 2005), pp. 436–446.
- [92] S. Domcke, R. Sinha, D. A. Levine, C. Sander, and N. Schultz. “Evaluating cell lines as tumour models by comparison of genomic profiles”. In: *Nat. Commun.* 4 (2013), p. 2126.
- [93] E. Durr, J. Yu, K. M. Krasinska, L. A. Carver, J. R. Yates III, J. E. Testa, P. Oh, and J. E. Schnitzer. “Direct proteomic mapping of the lung microvascular endothelial cell surface in vivo and in cell culture”. In: *Nat. Biotechnol.* 22.8 (Aug. 2004), pp. 985–992.

- [94] P. Carmeliet and R. K. Jain. “Angiogenesis in cancer and other diseases”. In: *Nature* 407.6801 (Sept. 2000), pp. 249–257.
- [95] J. R. Jackson, M. P. Seed, C. H. Kircher, D. A. Willoughby, and J. D. Winkler. “The codependence of angiogenesis and chronic inflammation”. In: *FASEB J.* 11.6 (May 1997), pp. 457–465.
- [96] I. E. Koutroubakis, G. Tsiolakidou, K. Karmiris, and E. A. Kouroumalis. “Role of angiogenesis in inflammatory bowel disease”. In: *Inflamm. Bowel Dis.* 12.6 (June 2006), pp. 515–523.
- [97] J. Medina, A. G. Arroyo, F. Sánchez-Madrid, and R. Moreno-Otero. “Angiogenesis in chronic inflammatory liver disease”. In: *Hepatology* 39.5 (May 2004), pp. 1185–1195.
- [98] X. Yang, P. Qiu, B. Chen, Y. Lin, Z. Zhou, R. Ge, H. Zou, J. Wang, and J. Wang. “KIAA1199 as a potential diagnostic biomarker of rheumatoid arthritis related to angiogenesis”. In: *Arthritis Res. Ther.* 17 (2015), p. 140.
- [99] A. E. Koch. “Review: angiogenesis: implications for rheumatoid arthritis”. In: *Arthritis Rheum.* 41.6 (June 1998), pp. 951–962.
- [100] C. X. Maracle and S. W. Tas. “Inhibitors of angiogenesis: ready for prime time?” In: *Best Pract. Res. Clin. Rheumatol.* 28.4 (Aug. 2014), pp. 637–649.
- [101] R. Heidenreich, M. Rocken, and K. Ghoreschi. “Angiogenesis drives psoriasis pathogenesis”. In: *Int. J. Exp. Pathol.* 90.3 (June 2009), pp. 232–248.
- [102] R. S. Sulaiman, H. D. Basavarajappa, and T. W. Corson. “Natural product inhibitors of ocular angiogenesis”. In: *Exp. Eye Res.* 129 (Dec. 2014), pp. 161–171.
- [103] T. A. Ciulla, A. G. Amador, and B. Zinman. “Diabetic retinopathy and diabetic macular edema: pathophysiology, screening, and novel therapies”. In: *Diabetes Care* 26.9 (Sept. 2003), pp. 2653–2664.
- [104] N. Ferrara and K. Alitalo. “Clinical applications of angiogenic growth factors and their inhibitors”. In: *Nat. Med.* 5.12 (Dec. 1999), pp. 1359–1364.
- [105] F. Woitek et al. “Intracoronary Cytoprotective Gene Therapy: A Study of VEGF-B167 in a Pre-Clinical Animal Model of Dilated Cardiomyopathy”. In: *J. Am. Coll. Cardiol.* 66.2 (July 2015), pp. 139–153.
- [106] F. J. Burrows and P. E. Thorpe. “Eradication of large solid tumors in mice with an immunotoxin directed against tumor vasculature”. In: *Proc. Natl. Acad. Sci. U.S.A.* 90.19 (Oct. 1993), pp. 8996–9000.
- [107] R. K. Jain. “A new target for tumor therapy”. In: *N. Engl. J. Med.* 360.25 (June 2009), pp. 2669–2671.
- [108] M. Mazzone et al. “Heterozygous deficiency of PHD2 restores tumor oxygenation and inhibits metastasis via endothelial normalization”. In: *Cell* 136.5 (Mar. 2009), pp. 839–851.

- [109] S. S. Chang, V. E. Reuter, W. D. W. Heston, N. H. Bander, L. S. Grauer, and P. B. Gaudin. “Five different anti-prostate-specific membrane antigen (PSMA) antibodies confirm PSMA expression in tumor-associated neovascu-
lature”. In: *Cancer Res.* 59.13 (July 1999), pp. 3192–3198.
- [110] N. Pandit-Taskar, J. A. O’Donoghue, C. R. Divgi, E. A. Wills, L. Schwartz, M. Gönen, P. Smith-Jones, N. H. Bander, H. I. Scher, S. M. Larson, and M. J. Morris. “Indium 111-labeled J591 anti-PSMA antibody for vascular targeted
imaging in progressive solid tumors”. In: *EJNMMI Res.* 5 (2015), p. 28.
- [111] C. Frantz, K. M. Stewart, and V. M. Weaver. “The extracellular matrix at a
glance”. In: *J. Cell. Sci.* 123.Pt 24 (Dec. 2010), pp. 4195–4200.
- [112] P. Lu, K. Takai, V. M. Weaver, and Z. Werb. “Extracellular matrix degra-
dation and remodeling in development and disease”. In: *Cold Spring Harb. Perspect. Biol.* 3.12 (Dec. 2011).
- [113] M. W. Pickup, J. K. Mouw, and V. M. Weaver. “The extracellular ma-
trix modulates the hallmarks of cancer”. In: *EMBO Rep.* 15.12 (Dec. 2014), pp. 1243–1253.
- [114] P. Lu, V. M. Weaver, and Z. Werb. “The extracellular matrix: a dynamic
niche in cancer progression”. In: *J. Cell Biol.* 196.4 (Feb. 2012), pp. 395–406.
- [115] B. C. Berk, K. Fujiwara, and S. Lehoux. “ECM remodeling in hypertensive
heart disease”. In: *J. Clin. Invest.* 117.3 (Mar. 2007), pp. 568–575.
- [116] H. Järveläinen, A. Sainio, M. Koulu, T. N. Wight, and R. Penttinen. “Extra-
cellular matrix molecules: potential targets in pharmacotherapy”. In: *Phar-
macol. Rev.* 61.2 (June 2009), pp. 198–223.
- [117] T. Hemmerle, S. Zraggen, M. Matasci, C. Halin, M. Detmar, and D. Neri. “Antibody-mediated delivery of interleukin 4 to the neo-vasculature reduces
chronic skin inflammation”. In: *J. Dermatol. Sci.* 76.2 (Nov. 2014), pp. 96–103.
- [118] F. Quattrone, A. M. Sanchez, M. Pannese, T. Hemmerle, P. Viganò, M. Candiani, F. Petraglia, D. Neri, and P. Panina-Bordignon. “The Targeted
Delivery of Interleukin 4 Inhibits Development of Endometriotic Lesions in a Mouse Model”. In: *Reprod. Sci.* (Apr. 2015).
- [119] C. C. Wu and J. R. Yates. “The application of mass spectrometry to mem-
brane proteomics”. In: *Nat. Biotechnol.* 21.3 (Mar. 2003), pp. 262–267.
- [120] A. O. Helbig, A. J. R. Heck, and M. Slijper. “Exploring the membrane
proteome—challenges and analytical strategies”. In: *J. Proteomics* 73.5 (Mar. 2010), pp. 868–878.
- [121] K.-C. Chou and D. W. Elrod. “Prediction of membrane protein types and
subcellular locations”. In: *Proteins* 34.1 (Jan. 1999), pp. 137–153.

- [122] B. Trzaskowski, D. Latek, S. Yuan, U. Ghoshdastider, A. Debinski, and S. Filipek. “Action of molecular switches in GPCRs—theoretical and experimental studies”. In: *Curr. Med. Chem.* 19.8 (2012), pp. 1090–1109.
- [123] M. S. Almén, K. J. V. Nordstrom, R. Fredriksson, and H. B. Schiöth. “Mapping the human membrane proteome: a majority of the human membrane proteins can be classified according to function and evolutionary origin”. In: *BMC Biol.* 7 (2009), p. 50.
- [124] E. Wallin and G. von Heijne. “Genome-wide analysis of integral membrane proteins from eubacterial, archaean, and eukaryotic organisms”. In: *Protein Sci.* 7.4 (Apr. 1998), pp. 1029–1038.
- [125] M. Ahram, Z. I. Litou, R. Fang, and G. Al-Tawallbeh. “Estimation of membrane proteins in the human proteome”. In: *In Silico Biol.* 6.5 (2006), pp. 379–386.
- [126] S. B. Scheurer, J. N. Rybak, C. Roesli, R. A. Brunisholz, F. Potthast, R. Schlapbach, D. Neri, and G. Elia. “Identification and relative quantification of membrane proteins by surface biotinylation and two-dimensional peptide mapping”. In: *Proteomics* 5.11 (July 2005), pp. 2718–2728.
- [127] A. L. Hopkins and C. R. Groom. “The druggable genome”. In: *Nat. Rev. Drug Discov.* 1.9 (Sept. 2002), pp. 727–730.
- [128] Z. Cheng, D. Garvin, A. Paguio, P. Stecha, K. Wood, and F. Fan. “Luciferase Reporter Assay System for Deciphering GPCR Pathways”. In: *Curr. Chem. Genomics* 4 (2010), pp. 84–91.
- [129] J. Drews. “Drug discovery: a historical perspective”. In: *Science* 287.5460 (Mar. 2000), pp. 1960–1964.
- [130] N. Nagaraj, J. R. Wisniewski, T. Geiger, J. Cox, M. Kircher, J. Kelso, S. Pääbo, and M. Mann. “Deep proteome and transcriptome mapping of a human cancer cell line”. In: *Mol. Syst. Biol.* 7 (2011), p. 548.
- [131] N. S. Madhukar, M. O. Warmoes, and J. W. Locasale. “Organization of enzyme concentration across the metabolic network in cancer cells”. In: *PLoS ONE* 10.1 (2015), e0117131.
- [132] B. A. Macher and T.-Y. Yen. “Proteins at membrane surfaces—a review of approaches”. In: *Mol. Biosyst.* 3.10 (Oct. 2007), pp. 705–713.
- [133] S. Tan, H. T. Tan, and M. C. M. Chung. “Membrane proteins and membrane proteomics”. In: *Proteomics* 8.19 (Oct. 2008), pp. 3924–3932.
- [134] D. K. Han, J. Eng, H. Zhou, and R. Aebersold. “Quantitative profiling of differentiation-induced microsomal proteins using isotope-coded affinity tags and mass spectrometry”. In: *Nat. Biotechnol.* 19.10 (Oct. 2001), pp. 946–951.
- [135] N. Nagaraj, A. Lu, M. Mann, and J. R. Wiśniewski. “Detergent-based but gel-free method allows identification of several hundred membrane proteins in single LC-MS runs”. In: *J. Proteome Res.* 7.11 (Nov. 2008), pp. 5028–5032.

- [136] J. Blonder, M. B. Goshe, R. J. Moore, L. Pasa-Tolic, C. D. Masselon, M. S. Lipton, and R. D. Smith. “Enrichment of integral membrane proteins for proteomic analysis using liquid chromatography-tandem mass spectrometry”. In: *J. Proteome Res.* 1.4 (2002), pp. 351–360.
- [137] M. P. Washburn, D. Wolters, and J. R. Yates III. “Large-scale analysis of the yeast proteome by multidimensional protein identification technology”. In: *Nat. Biotechnol.* 19.3 (Mar. 2001), pp. 242–247.
- [138] Y. G. Yeung, E. Nieves, R. H. Angeletti, and E. R. Stanley. “Removal of detergents from protein digests for mass spectrometry analysis”. In: *Anal. Biochem.* 382.2 (Nov. 2008), pp. 135–137.
- [139] D. Vuckovic, L. F. Dagley, A. W. Purcell, and A. Emili. “Membrane proteomics by high performance liquid chromatography-tandem mass spectrometry: Analytical approaches and challenges”. In: *Proteomics* 13.3-4 (Feb. 2013), pp. 404–423.
- [140] Y. Q. Yu, M. Gilar, P. J. Lee, E. S. Bouvier, and J. C. Gebler. “Enzyme-friendly, mass spectrometry-compatible surfactant for in-solution enzymatic digestion of proteins”. In: *Anal. Chem.* 75.21 (Nov. 2003), pp. 6023–6028.
- [141] L. A. Eichacker, B. Granvogl, O. Mirus, B. C. Muller, C. Miess, and E. Schleiff. “Hiding behind hydrophobicity. Transmembrane segments in mass spectrometry”. In: *J. Biol. Chem.* 279.49 (Dec. 2004), pp. 50915–50922.
- [142] B. A. van Montfort, M. K. Doeven, B. Canas, L. M. Veenhoff, B. Poolman, and G. T. Robillard. “Combined in-gel tryptic digestion and CNBr cleavage for the generation of peptide maps of an integral membrane protein with MALDI-TOF mass spectrometry”. In: *Biochim. Biophys. Acta* 1555.1-3 (Sept. 2002), pp. 111–115.
- [143] Y. Fujiki, A. L. Hubbard, S. Fowler, and P. B. Lazarow. “Isolation of intracellular membranes by means of sodium carbonate treatment: application to endoplasmic reticulum”. In: *J. Cell Biol.* 93.1 (Apr. 1982), pp. 97–102.
- [144] F. Elortza, S. Mohammed, J. Bunkenborg, L. J. Foster, T. S. Nühse, U. Brodbeck, S. C. Peck, and O. N. Jensen. “Modification-specific proteomics of plasma membrane proteins: identification and characterization of glycosylphosphatidylinositol-anchored proteins released upon phospholipase D treatment”. In: *J. Proteome Res.* 5.4 (Apr. 2006), pp. 935–943.
- [145] C. C. Wu, M. J. MacCoss, K. E. Howell, and J. R. Yates III. “A method for the comprehensive proteomic analysis of membrane proteins”. In: *Nat. Biotech.* 21 (May 2003), pp. 532–538.
- [146] S. Elschenbroich, Y. Kim, J. A. Medin, and T. Kislinger. “Isolation of cell surface proteins for mass spectrometry-based proteomics”. In: *Expert Rev. Proteomic* 7.1 (Feb. 2010), pp. 141–154.

- [147] W. Dormeyer, D. van Hoof, S. R. Braam, A. J. Heck, C. L. Mummery, and J. Krijgsveld. “Plasma membrane proteomics of human embryonic stem cells and human embryonal carcinoma cells”. In: *J. Proteome Res.* 7.7 (July 2008), pp. 2936–2951.
- [148] F. J. Burrows, E. J. Derbyshire, P. L. Tazzari, P. Amlot, A. F. Gazdar, S. W. King, M. Letarte, E. S. Vitetta, and P. E. Thorpe. “Up-regulation of endoglin on vascular endothelial cells in human solid tumors: implications for diagnosis and therapy”. In: *Clin. Cancer Res.* 1.12 (Dec. 1995), pp. 1623–1634.
- [149] H. J. Bühring, C. A. Müller, M. Letarte, A. Gougos, A. Saalmüller, A. J. van Agthoven, and F. W. Busch. “Endoglin is expressed on a subpopulation of immature erythroid cells of normal human bone marrow”. In: *Leukemia* 5.10 (Oct. 1991), pp. 841–847.
- [150] H. Liu, P. Moy, S. Kim, Y. Xia, A. Rajasekaran, V. Navarro, B. Knudsen, and N. H. Bander. “Monoclonal antibodies to the extracellular domain of prostate-specific membrane antigen also react with tumor vascular endothelium”. In: *Cancer Res.* 57.17 (Sept. 1997), pp. 3629–3634.
- [151] E. Balza, P. Castellani, A. Zijlstra, D. Neri, L. Zardi, and A. Siri. “Lack of specificity of endoglin expression for tumor blood vessels”. In: *Int. J. Cancer* 94.4 (Nov. 2001), pp. 579–585.
- [152] R. Pasqualini and E. Ruoslahti. “Organ targeting in vivo using phage display peptide libraries”. In: *Nature* 380.6572 (Mar. 1996), pp. 364–366.
- [153] W. Arap, R. Pasqualini, and E. Ruoslahti. “Cancer treatment by targeted drug delivery to tumor vasculature in a mouse model”. In: *Science* 279.5349 (Jan. 1998), pp. 377–380.
- [154] G. P. Smith. “Filamentous fusion phage: novel expression vectors that display cloned antigens on the virion surface”. In: *Science* 228.4705 (June 1985), pp. 1315–1317.
- [155] C. M. Hammers and J. R. Stanley. “Antibody phage display: technique and applications”. In: *J. Invest. Dermatol.* 134.2 (Feb. 2014), e17.
- [156] R. Pasqualini, E. Koivunen, R. Kain, J. Lahdenranta, M. Sakamoto, A. Stryhn, R. A. Ashmun, L. H. Shapiro, W. Arap, and E. Ruoslahti. “Aminopeptidase N is a receptor for tumor-homing peptides and a target for inhibiting angiogenesis”. In: *Cancer Res.* 60.3 (Feb. 2000), pp. 722–727.
- [157] P. Laakkonen, K. Porkka, J. A. Hoffman, and E. Ruoslahti. “A tumor-homing peptide with a targeting specificity related to lymphatic vessels”. In: *Nat. Med.* 8.7 (July 2002), pp. 751–755.
- [158] P. Laakkonen, L. Zhang, and E. Ruoslahti. “Peptide targeting of tumor lymph vessels”. In: *Ann. N. Y. Acad. Sci.* 1131 (2008), pp. 37–43.

- [159] J. M. J. Herbert, D. Stekel, S. Sanderson, V. L. Heath, and R. Bicknell. “A novel method of differential gene expression analysis using multiple cDNA libraries applied to the identification of tumour endothelial genes”. In: *BMC Genomics* 9 (2008), p. 153.
- [160] V. Strassberger, T. Fugmann, D. Neri, and C. Rösli. “Chemical proteomic and bioinformatic strategies for the identification and quantification of vascular antigens in cancer”. In: *J. Proteomics* 73.10 (Sept. 2010), pp. 1954–1973.
- [161] X. Wei and L. Li. “Comparative glycoproteomics: approaches and applications”. In: *Brief. Funct. Genomic. Proteomic.* 8.2 (Mar. 2009), pp. 104–113.
- [162] M. Yanagi, Y. Aoyagi, T. Suda, Y. Mita, and H. Asakura. “N-Acetylglucosaminyltransferase V as a possible aid for the evaluation of tumor invasiveness in patients with hepatocellular carcinoma”. In: *J. Gastroenterol. Hepatol.* 16.11 (Nov. 2001), pp. 1282–1289.
- [163] S. F. Siddiqui, J. Pawelek, T. Handerson, C.-Y. Lin, R. B. Dickson, D. L. Rimm, and R. L. Camp. “Coexpression of beta1,6-N-acetylglucosaminyltransferase V glycoprotein substrates defines aggressive breast cancers with poor outcome”. In: *Cancer Epidemiol. Biomarkers Prev.* 14.11 (Nov. 2005), pp. 2517–2523.
- [164] R. Lotan and G. L. Nicolson. “Purification of cell membrane glycoproteins by lectin affinity chromatography”. In: *Biochim. Biophys. Acta* 559.4 (Dec. 1979), pp. 329–376.
- [165] R. D. Cummings and S. Kornfeld. “Fractionation of asparagine-linked oligosaccharides by serial lectin-Agarose affinity chromatography. A rapid, sensitive, and specific technique”. In: *J. Biol. Chem.* 257.19 (Oct. 1982), pp. 11235–11240.
- [166] R. Qiu and F. E. Regnier. “Comparative glycoproteomics of N-linked complex-type glycoforms containing sialic acid in human serum”. In: *Anal. Chem.* 77.22 (Nov. 2005), pp. 7225–7231.
- [167] H. Kaji, H. Saito, Y. Yamauchi, T. Shinkawa, M. Taoka, J. Hirabayashi, K. Kasai, N. Takahashi, and T. Isobe. “Lectin affinity capture, isotope-coded tagging and mass spectrometry to identify N-linked glycoproteins”. In: *Nat. Biotechnol.* 21.6 (June 2003), pp. 667–672.
- [168] K. Sparbier, T. Wenzel, and M. Kostrzewa. “Exploring the binding profiles of ConA, boronic acid and WGA by MALDI-TOF/TOF MS and magnetic particles”. In: *J. Chromatogr. B* 840.1 (Aug. 2006), pp. 29–36.
- [169] K. Vosseller, K. C. Hansen, R. J. Chalkley, J. C. Trinidad, L. Wells, G. W. Hart, and A. L. Burlingame. “Quantitative analysis of both protein expression and serine / threonine post-translational modifications through stable isotope labeling with dithiothreitol”. In: *Proteomics* 5.2 (Feb. 2005), pp. 388–398.

- [170] N. Khidekel, S. B. Ficarro, P. M. Clark, M. C. Bryan, D. L. Swaney, J. E. Rexach, Y. E. Sun, J. J. Coon, E. C. Peters, and L. C. Hsieh-Wilson. “Probing the dynamics of O-GlcNAc glycosylation in the brain using quantitative proteomics”. In: *Nat. Chem. Biol.* 3.6 (June 2007), pp. 339–348.
- [171] H. Zhang, X.-J. Li, D. B. Martin, and R. Aebersold. “Identification and quantification of N-linked glycoproteins using hydrazide chemistry, stable isotope labeling and mass spectrometry”. In: *Nat. Biotechnol.* 21.6 (June 2003), pp. 660–666.
- [172] B. Wollscheid, D. Bausch-Fluck, C. Henderson, R. O’Brien, M. Bibel, R. Schiess, R. Aebersold, and J. D. Watts. “Mass-spectrometric identification and relative quantification of N-linked cell surface glycoproteins”. In: *Nat. Biotechnol.* 27.4 (Apr. 2009), pp. 378–386.
- [173] C. G. Gahmberg and M. Tolvanen. “Why mammalian cell surface proteins are glycoproteins”. In: *Trends Biochem. Sci.* 21.8 (Aug. 1996), pp. 308–311.
- [174] L. K. Chaney and B. S. Jacobson. “Coating cells with colloidal silica for high yield isolation of plasma membrane sheets and identification of transmembrane proteins”. In: *J. Biol. Chem.* 258.16 (Aug. 1983), pp. 10062–10072.
- [175] B. S. Jacobson, J. E. Schnitzer, M. McCaffery, and G. E. Palade. “Isolation and partial characterization of the luminal plasmalemma of microvascular endothelium from rat lungs”. In: *Eur. J. Cell Biol.* 58.2 (Aug. 1992), pp. 296–306.
- [176] A. M. Rahbar and C. Fenselau. “Integration of Jacobson’s pellicle method into proteomic strategies for plasma membrane proteins”. In: *J. Proteome Res.* 3.6 (2004), pp. 1267–1277.
- [177] A. M. Rahbar and C. Fenselau. “Unbiased examination of changes in plasma membrane proteins in drug resistant cancer cells”. In: *J. Proteome Res.* 4.6 (2005), pp. 2148–2153.
- [178] S. Arjunan, M. Reinartz, B. Emde, K. Zanger, and J. Schrader. “Limitations of the colloidal silica method in mapping the endothelial plasma membrane proteome of the mouse heart”. In: *Cell Biochem. Biophys.* 53.3 (2009), pp. 135–143.
- [179] A. W. Brändli, R. G. Parton, and K. Simons. “Transcytosis in MDCK cells: identification of glycoproteins transported bidirectionally between both plasma membrane domains”. In: *J. Cell Biol.* 111.6 Pt. 2 (Dec. 1990), pp. 2909–2921.
- [180] M. P. Weekes, R. Antrobus, J. R. Lill, L. M. Duncan, S. Hör, and P. J. Lehner. “Comparative analysis of techniques to purify plasma membrane proteins”. In: *J. Biomol. Tech.* 21.3 (Sept. 2010), pp. 108–115.
- [181] Y. Zhao, W. Zhang, Y. Kho, and Y. Zhao. “Proteomic analysis of integral plasma membrane proteins”. In: *Anal. Chem.* 76.7 (Apr. 2004), pp. 1817–1823.

- [182] M. J. Peirce, R. Wait, S. Begum, J. Saklatvala, and A. P. Cope. “Expression profiling of lymphocyte plasma membrane proteins”. In: *Mol. Cell. Proteomics* 3.1 (Jan. 2004), pp. 56–65.
- [183] C. Rösli, B. Borgia, C. Schliemann, M. Gunthert, H. Wunderli-Allenspach, R. Giavazzi, and D. Neri. “Comparative analysis of the membrane proteome of closely related metastatic and nonmetastatic tumor cells”. In: *Cancer Res.* 69.13 (July 2009), pp. 5406–5414.
- [184] J.-N. Rybak, A. Ettorre, B. Kaissling, R. Giavazzi, D. Neri, and G. Elia. “In vivo protein biotinylation for identification of organ-specific antigens accessible from the vasculature”. In: *Nat. Methods* 2.4 (Apr. 2005), pp. 291–298.
- [185] C. Rösli, D. Neri, and J. N. Rybak. “In vivo protein biotinylation and sample preparation for the proteomic identification of organ- and disease-specific antigens accessible from the vasculature”. In: *Nat. Protoc.* 1.1 (2006), pp. 192–199.
- [186] J. N. Rybak, C. Rösli, M. Kaspar, A. Villa, and D. Neri. “The extra-domain A of fibronectin is a vascular marker of solid tumors and metastases”. In: *Cancer Res.* 67.22 (Nov. 2007), pp. 10948–10957.
- [187] B. Borgia, C. Rösli, T. Fugmann, C. Schliemann, M. Cesca, D. Neri, and R. Giavazzi. “A proteomic approach for the identification of vascular markers of liver metastasis”. In: *Cancer Res.* 70.1 (Jan. 2010), pp. 309–318.
- [188] C. Schliemann, C. Roesli, H. Kamada, B. Borgia, T. Fugmann, W. Klapper, and D. Neri. “In vivo biotinylation of the vasculature in B-cell lymphoma identifies BST-2 as a target for antibody-based therapy”. In: *Blood* 115.3 (Jan. 2010), pp. 736–744.
- [189] V. Castronovo, D. Waltregny, P. Kischel, C. Rösli, G. Elia, J.-N. Rybak, and D. Neri. “A chemical proteomics approach for the identification of accessible antigens expressed in human kidney cancer”. In: *Mol. Cell. Proteomics* 5.11 (Nov. 2006), pp. 2083–2091.
- [190] P. Conrotto, C. Rösli, J. Rybak, P. Kischel, D. Waltregny, D. Neri, and V. Castronovo. “Identification of new accessible tumor antigens in human colon cancer by ex vivo protein biotinylation and comparative mass spectrometry analysis”. In: *Int. J. Cancer* 123.12 (Dec. 2008), pp. 2856–2864.
- [191] G. Elia. “Biotinylation reagents for the study of cell surface proteins”. In: *Proteomics* 8.19 (Oct. 2008), pp. 4012–4024.
- [192] J. R. Yates, C. I. Ruse, and A. Nakorchevsky. “Proteomics by mass spectrometry: approaches, advances, and applications”. In: *Annu. Rev. Biomed. Eng.* 11 (2009), pp. 49–79.
- [193] F. Meissner and M. Mann. “Quantitative shotgun proteomics: considerations for a high-quality workflow in immunology”. In: *Nat. Immunol.* 15.2 (Feb. 2014), pp. 112–117.

- [194] J. C. Tran, L. Zamdborg, D. R. Ahlf, J. E. Lee, A. D. Catherman, K. R. Durbin, J. D. Tipton, A. Vellaichamy, J. F. Kellie, M. Li, C. Wu, S. M. Sweet, B. P. Early, N. Siuti, R. D. LeDuc, P. D. Compton, P. M. Thomas, and N. L. Kelleher. "Mapping intact protein isoforms in discovery mode using top-down proteomics". In: *Nature* 480.7376 (Dec. 2011), pp. 254–258.
- [195] A. Armirotti and G. Damonte. "Achievements and perspectives of top-down proteomics". In: *Proteomics* 10.20 (Oct. 2010), pp. 3566–3576.
- [196] N. L. Kelleher. "Top-down proteomics". In: *Anal. Chem.* 76.11 (June 2004), 197A–203A.
- [197] K. M. Millea, I. S. Krull, S. A. Cohen, J. C. Gebler, and S. J. Berger. "Integration of multidimensional chromatographic protein separations with a combined "top-down" and "bottom-up" proteomic strategy". In: *J. Proteome Res.* 5.1 (Jan. 2006), pp. 135–146.
- [198] T. Fröhlich and G. J. Arnold. "Proteome research based on modern liquid chromatography–tandem mass spectrometry: separation, identification and quantification". In: *J. Neural. Transm.* 113.8 (Aug. 2006), pp. 973–994.
- [199] C. May, F. Broseron, P. Chartowski, C. Schumbrutzki, B. Schoenebeck, and K. Marcus. "Instruments and methods in proteomics". In: *Data Mining in Proteomics*. Springer, 2011, pp. 3–26.
- [200] Y. Shen, R. Zhang, R. J. Moore, J. Kim, T. O. Metz, K. K. Hixson, R. Zhao, E. A. Livesay, H. R. Udseth, and R. D. Smith. "Automated 20 kpsi RPLC-MS and MS/MS with chromatographic peak capacities of 1000-1500 and capabilities in proteomics and metabolomics". In: *Anal. Chem.* 77.10 (May 2005), pp. 3090–3100.
- [201] A. P. Lothrop, M. P. Torres, and S. M. Fuchs. "Deciphering post-translational modification codes". In: *FEBS Lett.* 587.8 (Apr. 2013), pp. 1247–1257.
- [202] X. Shen, R. Young, J. M. Canty, and J. Qu. "Quantitative proteomics in cardiovascular research: global and targeted strategies". In: *Proteomics Clin. Appl.* 8.7-8 (Aug. 2014), pp. 488–505.
- [203] G. A. Khoury, R. C. Baliban, and C. A. Floudas. "Proteome-wide post-translational modification statistics: frequency analysis and curation of the swiss-prot database". In: *Sci. Rep.* 1 (Sept. 2011).
- [204] K. Engholm-Keller and M. R. Larsen. "Technologies and challenges in large-scale phosphoproteomics". In: *Proteomics* 13.6 (Mar. 2013), pp. 910–931.
- [205] N. D. Udeshi, P. Mertins, T. Svinkina, and S. A. Carr. "Large-scale identification of ubiquitination sites by mass spectrometry". In: *Nat. Protoc.* 8.10 (Oct. 2013), pp. 1950–1960.
- [206] R. Aebersold and M. Mann. "Mass spectrometry-based proteomics". In: *Nature* 422.6928 (Mar. 2003), pp. 198–207.

- [207] J. B. Fenn, M. Mann, C. K. Meng, S. F. Wong, and C. M. Whitehouse. "Electrospray ionization for mass spectrometry of large biomolecules". In: *Science* 246.4926 (Oct. 1989), pp. 64–71.
- [208] M. Karas and F. Hillenkamp. "Laser desorption ionization of proteins with molecular masses exceeding 10,000 daltons". In: *Anal. Chem.* 60.20 (Oct. 1988), pp. 2299–2301.
- [209] K. Tanaka, H. Waki, Y. Ido, S. Akita, Y. Yoshida, T. Yoshida, and T. Matsuo. "Protein and polymer analyses up to m/z 100 000 by laser ionization time-of-flight mass spectrometry". In: *Rapid Commun. Mass Sp.* 2.8 (1988), pp. 151–153.
- [210] M. Wilm. "Principles of electrospray ionization". In: *Mol. Cell. Proteomics* 10.7 (July 2011), p. M111.009407.
- [211] R. Cramer. "MALDI MS". In: *Methods Mol. Biol.* 564 (2009), pp. 85–103.
- [212] U. Bahr, M. Karas, and F. Hillenkamp. "Analysis of biopolymers by matrix-assisted laser desorption/ionization (MALDI) mass spectrometry". In: *Fresenius J. Anal. Chem.* 348.12 (Dec. 1994), pp. 783–791.
- [213] C. Eriksson, N. Masaki, I. Yao, T. Hayasaka, and M. Setou. "MALDI Imaging Mass Spectrometry-A Mini Review of Methods and Recent Developments". In: *Mass. Spectrom.* 2.Spec. Iss. (2013), S0022.
- [214] A. Walch, S. Rauser, S.-O. Deininger, and H. Höfler. "MALDI imaging mass spectrometry for direct tissue analysis: a new frontier for molecular histology". In: *Histochem. Cell Biol.* 130.3 (Sept. 2008), pp. 421–434.
- [215] G. L. Glush and D. J. Burinsky. "Hybrid mass spectrometers for tandem mass spectrometry". In: *J. Am. Soc. Mass Spectrom.* 19.2 (Feb. 2008), pp. 161–172.
- [216] X. Han, A. Aslanian, and J. R. Yates. "Mass spectrometry for proteomics". In: *Curr. Opin. Chem. Biol.* 12.5 (Oct. 2008), pp. 483–490.
- [217] M. Scigelova, M. Hornshaw, A. Giannakopoulos, and A. Makarov. "Fourier transform mass spectrometry". In: *Mol. Cell. Proteomics* 10.7 (July 2011), p. M111.009431.
- [218] A. Makarov and M. Scigelova. "Coupling liquid chromatography to Orbitrap mass spectrometry". In: *J. Chromatogr. A* 1217.25 (June 2010), pp. 3938–3945.
- [219] J. K. Eng, B. C. Searle, K. R. Clauser, and D. L. Tabb. "A face in the crowd: recognizing peptides through database search". In: *Mol. Cell. Proteomics* 10.11 (Nov. 2011), R111.009522.
- [220] Y. Liu, R. Hüttenhain, B. Collins, and R. Aebersold. "Mass spectrometric protein maps for biomarker discovery and clinical research". In: *Expert Rev. Mol. Diagn.* 13.8 (Nov. 2013), pp. 811–825.

- [221] J. S. Cottrell. “Protein identification using MS/MS data”. In: *J. Proteomics* 74.10 (Sept. 2011), pp. 1842–1851.
- [222] J. Seidler, N. Zinn, M. E. Boehm, and W. D. Lehmann. “De novo sequencing of peptides by MS/MS”. In: *Proteomics* 10.4 (Feb. 2010), pp. 634–649.
- [223] M. H. Elliott, D. S. Smith, C. E. Parker, and C. Borchers. “Current trends in quantitative proteomics”. In: *J. Mass Spectrom.* 44.12 (Dec. 2009), pp. 1637–1660.
- [224] T. Geiger, J. Cox, P. Ostasiewicz, J. R. Wisniewski, and M. Mann. “Super-SILAC mix for quantitative proteomics of human tumor tissue”. In: *Nat. Methods* 7.5 (May 2010), pp. 383–385.
- [225] J. Krijgsveld, R. F. Ketting, T. Mahmoudi, J. Johansen, M. Artal-Sanz, C. P. Verrijzer, R. H. Plasterk, and A. J. Heck. “Metabolic labeling of *C. elegans* and *D. melanogaster* for quantitative proteomics”. In: *Nat. Biotechnol.* 21.8 (Aug. 2003), pp. 927–931.
- [226] S. Zanivan, M. Krueger, and M. Mann. “In vivo quantitative proteomics: the SILAC mouse”. In: *Methods Mol. Biol.* 757 (2012), pp. 435–450.
- [227] K. J. Reynolds and C. Fenselau. “Quantitative protein analysis using proteolytic [18O]water labeling”. In: *Curr. Protoc. Protein Sci.* Chapter 23 (Feb. 2004), pp. 4.1–4.8.
- [228] S. P. Gygi, B. Rist, S. A. Gerber, F. Turecek, M. H. Gelb, and R. Aebersold. “Quantitative analysis of complex protein mixtures using isotope-coded affinity tags”. In: *Nat. Biotechnol.* 17.10 (Oct. 1999), pp. 994–999.
- [229] A. Schmidt, J. Kellermann, and F. Lottspeich. “A novel strategy for quantitative proteomics using isotope-coded protein labels”. In: *Proteomics* 5.1 (Jan. 2005), pp. 4–15.
- [230] M. B. Goshe, T. P. Conrads, E. A. Panisko, N. H. Angell, T. D. Veenstra, and R. D. Smith. “Phosphoprotein isotope-coded affinity tag approach for isolating and quantitating phosphopeptides in proteome-wide analyses”. In: *Anal. Chem.* 73.11 (June 2001), pp. 2578–2586.
- [231] C. Evans, J. Noirel, S. Y. Ow, M. Salim, A. G. Pereira-Medrano, N. Couto, J. Pandhal, D. Smith, T. K. Pham, E. Karunakaran, X. Zou, C. A. Biggs, and P. C. Wright. “An insight into iTRAQ: where do we stand now?” In: *Anal. Bioanal. Chem.* 404.4 (Sept. 2012), pp. 1011–1027.
- [232] D. Phanstiel, R. Unwin, G. C. McAlister, and J. J. Coon. “Peptide quantification using 8-plex isobaric tags and electron transfer dissociation tandem mass spectrometry”. In: *Anal. Chem.* 81.4 (Feb. 2009), pp. 1693–1698.
- [233] D. Zeng and S. Li. “Improved CILAT reagents for quantitative proteomics”. In: *Bioorg. Med. Chem. Lett.* 19.7 (Apr. 2009), pp. 2059–2061.

- [234] A. S. Benk and C. Rösli. “Label-free quantification using MALDI mass spectrometry: considerations and perspectives”. In: *Anal. Bioanal. Chem.* 404.4 (Sept. 2012), pp. 1039–1056.
- [235] J. Rappsilber, U. Ryder, A. I. Lamond, and M. Mann. “Large-scale proteomic analysis of the human spliceosome”. In: *Genome Res.* 12.8 (Aug. 2002), pp. 1231–1245.
- [236] Y. Ishihama, Y. Oda, T. Tabata, T. Sato, T. Nagasu, J. Rappsilber, and M. Mann. “Exponentially modified protein abundance index (emPAI) for estimation of absolute protein amount in proteomics by the number of sequenced peptides per protein”. In: *Mol. Cell. Proteomics* 4.9 (Sept. 2005), pp. 1265–1272.
- [237] K. A. Neilson, N. A. Ali, S. Muralidharan, M. Mirzaei, M. Mariani, G. As-sadourian, A. Lee, S. C. van Sluyter, and P. A. Haynes. “Less label, more free: approaches in label-free quantitative mass spectrometry”. In: *Proteomics* 11.4 (Feb. 2011), pp. 535–553.
- [238] B. L. Zybailov, L. Florens, and M. P. Washburn. “Quantitative shotgun proteomics using a protease with broad specificity and normalized spectral abundance factors”. In: *Mol. Biosyst.* 3.5 (May 2007), pp. 354–360.
- [239] J. C. Braisted, S. Kuntumalla, C. Vogel, E. M. Marcotte, A. R. Rodrigues, R. Wang, S. T. Huang, E. S. Ferlanti, A. I. Saeed, R. D. Fleischmann, S. N. Peterson, and R. Pieper. “The APEX Quantitative Proteomics Tool: generating protein quantitation estimates from LC-MS/MS proteomics results”. In: *BMC Bioinformatics* 9 (2008), p. 529.
- [240] M. Bantscheff, S. Lemeer, M. M. Savitski, and B. Kuster. “Quantitative mass spectrometry in proteomics: critical review update from 2007 to the present”. In: *Anal. Bioanal. Chem.* 404.4 (Sept. 2012), pp. 939–965.
- [241] D. Chelius and P. V. Bondarenko. “Quantitative profiling of proteins in complex mixtures using liquid chromatography and mass spectrometry”. In: *J. Proteome Res.* 1.4 (2002), pp. 317–323.
- [242] W. Zhu, J. W. Smith, and C.-M. Huang. “Mass spectrometry-based label-free quantitative proteomics”. In: *J. Biomed. Biotechnol.* 2010 (2010), p. 840518.
- [243] L. J. E. Goeminne, A. Argentini, L. Martens, and L. Clement. “Summarization vs Peptide-Based Models in Label-Free Quantitative Proteomics: Performance, Pitfalls, and Data Analysis Guidelines”. In: *J. Proteome Res.* 14.6 (June 2015), pp. 2457–2465.
- [244] K. P. Law and Y. P. Lim. “Recent advances in mass spectrometry: data independent analysis and hyper reaction monitoring”. In: *Expert Rev. Proteomics* 10.6 (Dec. 2013), pp. 551–566.

- [245] H. L. Röst, G. Rosenberger, P. Navarro, L. Gillet, S. M. Miladinović, O. T. Schubert, W. Wolski, B. C. Collins, J. Malmström, L. Malmström, and R. Aebersold. “OpenSWATH enables automated, targeted analysis of data-independent acquisition MS data”. In: *Nat. Biotechnol.* 32.3 (Mar. 2014), pp. 219–223.
- [246] J. D. Egertson, A. Kuehn, G. E. Merrihew, N. W. Bateman, B. X. MacLean, Y. S. Ting, J. D. Canterbury, D. M. Marsh, M. Kellmann, V. Zabrouskov, C. C. Wu, and M. J. MacCoss. “Multiplexed MS/MS for improved data-independent acquisition”. In: *Nat. Methods* 10.8 (Aug. 2013), pp. 744–746.
- [247] D. M. Cox, M. Jarvis, E. McClure, A. M. Taylor, and A. Morla. “Impact of novel Accurate Mass MS/MS ALL acquisition and processing techniques on forensic toxicological screening”. In: *Toxicologie Analytique et Clinique* 27.2 (2015), S52–S53.
- [248] A. S. Hebert, A. L. Richards, D. J. Bailey, A. Ulbrich, E. E. Coughlin, M. S. Westphall, and J. J. Coon. “The one hour yeast proteome”. In: *Mol. Cell. Proteomics* 13.1 (Jan. 2014), pp. 339–347.
- [249] U. Distler, J. Kuharev, P. Navarro, Y. Levin, H. Schild, and S. Tenzer. “Drift time-specific collision energies enable deep-coverage data-independent acquisition proteomics”. In: *Nat. Methods* 11.2 (Feb. 2014), pp. 167–170.
- [250] S. Surinova, R. Schiess, R. Hüttenhain, F. Cerciello, B. Wollscheid, and R. Aebersold. “On the development of plasma protein biomarkers”. In: *J. Proteome Res.* 10.1 (Jan. 2011), pp. 5–16.
- [251] D. J. Hunter, E. Losina, A. Guermazi, D. Burstein, M. N. Lasserre, and V. Kraus. “A pathway and approach to biomarker validation and qualification for osteoarthritis clinical trials”. In: *Curr. Drug Targets* 11.5 (May 2010), pp. 536–545.
- [252] G. Elia, T. Fugmann, and D. Neri. “From target discovery to clinical trials with armed antibody products”. In: *J. Proteomics* 107 (July 2014), pp. 50–55.
- [253] N. Rifai, M. A. Gillette, and S. A. Carr. “Protein biomarker discovery and validation: the long and uncertain path to clinical utility”. In: *Nat. Biotechnol.* 24.8 (2006), pp. 971–983.
- [254] D. S. Kirkpatrick, S. A. Gerber, and S. P. Gygi. “The absolute quantification strategy: a general procedure for the quantification of proteins and post-translational modifications”. In: *Methods* 35.3 (Mar. 2005), pp. 265–273.
- [255] N. L. Anderson, N. G. Anderson, L. R. Haines, D. B. Hardie, R. W. Olafson, and T. W. Pearson. “Mass spectrometric quantitation of peptides and proteins using Stable Isotope Standards and Capture by Anti-Peptide Antibodies (SISCAPA)”. In: *J. Proteome Res.* 3.2 (2004), pp. 235–244.

- [256] J. K. Scott and G. P. Smith. “Searching for peptide ligands with an epitope library”. In: *Science* 249.4967 (July 1990), pp. 386–390.
- [257] R. Frank. “The SPOT-synthesis technique. Synthetic peptide arrays on membrane supports—principles and applications”. In: *J. Immunol. Methods* 267.1 (Sept. 2002), pp. 13–26.
- [258] A. V. Terskikh, J.-M. Le Doussal, R. Crameri, I. Fisch, J.-P. Mach, and A. V. Kajava. “‘Peptabody’: a new type of high avidity binding protein”. In: *Proc. Natl. Acad. Sci. U.S.A.* 94.5 (Mar. 1997), pp. 1663–1668.
- [259] M. Houimel, P. Schneider, A. Terskikh, and J. P. Mach. “Selection of peptides and synthesis of pentameric peptabody molecules reacting specifically with ErbB-2 receptor”. In: *Int. J. Cancer* 92.5 (June 2001), pp. 748–755.
- [260] A. D. Ellington and J. W. Szostak. “Selection in vitro of single-stranded DNA molecules that fold into specific ligand-binding structures”. In: *Nature* 355.6363 (Feb. 1992), pp. 850–852.
- [261] D. A. Daniels, H. Chen, B. J. Hicke, K. M. Swiderek, and L. Gold. “A tenascin-C aptamer identified by tumor cell SELEX: systematic evolution of ligands by exponential enrichment”. In: *Proc. Natl. Acad. Sci. U.S.A.* 100.26 (Dec. 2003), pp. 15416–15421.
- [262] S. Helmling, C. Maasch, D. Eulberg, K. Buchner, W. Schröder, C. Lange, S. Vonhoff, B. Wlotzka, M. H. Tschöp, S. Rosewicz, and S. Klussmann. “Inhibition of ghrelin action in vitro and in vivo by an RNA-Spiegelmer”. In: *Proc. Natl. Acad. Sci. U.S.A.* 101.36 (Sept. 2004), pp. 13174–13179.
- [263] Y. Zhang, H. Hong, and W. Cai. “Tumor-targeted drug delivery with aptamers”. In: *Curr. Med. Chem.* 18.27 (2011), pp. 4185–4194.
- [264] S. L. Schreiber. “Organic synthesis toward small-molecule probes and drugs”. In: *Proc. Natl. Acad. Sci. U.S.A.* 108.17 (Apr. 2011), pp. 6699–6702.
- [265] O. Ramström and J.-M. Lehn. “Drug discovery by dynamic combinatorial libraries”. In: *Nat. Rev. Drug Discov.* 1.1 (Jan. 2002), pp. 26–36.
- [266] S. Melkko, J. Scheuermann, C. E. Dumelin, and D. Neri. “Encoded self-assembling chemical libraries”. In: *Nat. Biotechnol.* 22.5 (May 2004), pp. 568–574.
- [267] G. Casi and D. Neri. “Antibody-drug conjugates and small molecule-drug conjugates: opportunities and challenges for the development of selective anti-cancer cytotoxic agents”. In: *J. Med. Chem.* (June 2015).
- [268] R. T. Morris, R. N. Joyrich, R. W. Naumann, N. P. Shah, A. H. Maurer, H. W. Strauss, J. M. Uszler, J. T. Symanowski, P. R. Ellis, and W. A. Harb. “Phase II study of treatment of advanced ovarian cancer with folate-receptor-targeted therapeutic (vintafolide) and companion SPECT-based imaging agent (99m Tc-etarfolatide)”. In: *Ann. Oncol.* 25.4 (Apr. 2014), pp. 852–858.

- [269] C. Spiess, M. Merchant, A. Huang, Z. Zheng, N. Y. Yang, J. Peng, D. Ellerman, W. Shatz, D. Reilly, D. G. Yansura, and J. M. Scheer. “Bispecific antibodies with natural architecture produced by co-culture of bacteria expressing two distinct half-antibodies”. In: *Nat. Biotechnol.* 31.8 (Aug. 2013), pp. 753–758.
- [270] G. Köhler and C. Milstein. “Continuous cultures of fused cells secreting antibody of predefined specificity”. In: *Nature* 256.5517 (Aug. 1975), pp. 495–497.
- [271] G. Winter, A. D. Griffiths, R. E. Hawkins, and H. R. Hoogenboom. “Making antibodies by phage display technology”. In: *Annu. Rev. Immunol.* 12 (1994), pp. 433–455.
- [272] C. Schaffitzel, J. Hanes, L. Jermutus, and A. Plückthun. “Ribosome display: an in vitro method for selection and evolution of antibodies from libraries”. In: *J. Immunol. Methods* 231.1-2 (Dec. 1999), pp. 119–135.
- [273] L. Giovannoni, F. Viti, L. Zardi, and D. Neri. “Isolation of anti-angiogenesis antibodies from a large combinatorial repertoire by colony filter screening”. In: *Nucleic Acids Res.* 29.5 (Mar. 2001), E27.
- [274] G. J. Weiner. “Building better monoclonal antibody-based therapeutics”. In: *Nat. Rev. Cancer* 15.6 (June 2015), pp. 361–370.
- [275] M. Steiner and D. Neri. “Antibody-radionuclide conjugates for cancer therapy: historical considerations and new trends”. In: *Clin. Cancer Res.* 17.20 (Oct. 2011), pp. 6406–6416.
- [276] N. Pasche and D. Neri. “Immunocytokines: a novel class of potent armed antibodies”. In: *Drug Discov. Today* 17.11-12 (June 2012), pp. 583–590.
- [277] E. Perrino, M. Steiner, N. Krall, G. J. L. Bernardes, F. Pretto, G. Casi, and D. Neri. “Curative properties of noninternalizing antibody-drug conjugates based on maytansinoids”. In: *Cancer Res.* 74.9 (May 2014), pp. 2569–2578.
- [278] A. Wagh and B. Law. “Methods for conjugating antibodies to nanocarriers”. In: *Methods Mol. Biol.* 1045 (2013), pp. 249–266.
- [279] B. Nolting. “Linker technologies for antibody-drug conjugates”. In: *Methods Mol. Biol.* 1045 (2013), pp. 71–100.
- [280] T. Geiger, A. Velic, B. Macek, E. Lundberg, C. Kampf, N. Nagaraj, M. Uhlen, J. Cox, and M. Mann. “Initial quantitative proteomic map of 28 mouse tissues using the SILAC mouse”. In: *Mol. Cell. Proteomics* 12.6 (June 2013), pp. 1709–1722.
- [281] D. Wessel and U. I. Flügge. “A method for the quantitative recovery of protein in dilute solution in the presence of detergents and lipids”. In: *Anal. Biochem.* 138.1 (Apr. 1984), pp. 141–143.
- [282] P. A. Kerner. “MS_QBAT—A Software Suite for LC-MS Protein Quantification”. PhD thesis. University of Heidelberg, Germany, 2015.

- [283] B. Zhang, S. Kirov, and J. Snoddy. “WebGestalt: an integrated system for exploring gene sets in various biological contexts”. In: *Nucleic Acids Res.* 33 (July 2005), pp. 741–748.
- [284] J. Wang, D. Duncan, Z. Shi, and B. Zhang. “WEB-based GENE SeT AnaLysis Toolkit (WebGestalt): update 2013”. In: *Nucleic Acids Res.* 41 (July 2013), pp. 77–83.
- [285] J. Gould. *GENE-E*. URL: <http://www.broadinstitute.org/cancer/software/GENE-E/> (visited on 06/01/2015).
- [286] M. Wang, M. Weiss, M. Simonovic, G. Haertinger, S. P. Schimpf, M. O. Hengartner, and C. von Mering. “PaxDb, a database of protein abundance averages across all three domains of life”. In: *Mol. Cell. Proteomics* 11.8 (Aug. 2012), pp. 492–500.
- [287] M. Wang, C. J. Herrmann, M. Simonovic, D. Szklarczyk, and C. von Mering. “Version 4.0 of PaxDb: Protein abundance data, integrated across model organisms, tissues, and cell-lines”. In: *Proteomics* (Mar. 2015).
- [288] E. Gasteiger, C. Hoogland, A. Gattiker, S. Duvaud, M. R. Wilkins, R. D. Appel, and A. Bairoch. “Protein Identification and Analysis Tools on the ExPASy Server”. In: *The Proteomics Protocols Handbook*. Ed. by J. M. Walker. Humana Press, 2005, pp. 571–607.
- [289] A. M. Scott, J. D. Wolchok, and O. L. J. “Antibody therapy of cancer”. In: *Nat. Rev. Cancer* 12.4 (Apr. 2012), pp. 278–287.
- [290] Y. Z. Zheng and L. J. Foster. “Biochemical and proteomic approaches for the study of membrane microdomains”. In: *J. Proteomics* 72.1 (Feb. 2009), pp. 12–22.
- [291] N. M. Green. “Avidin. 3. The nature of the Biotin-binding site”. In: *Biochem. J.* 89 (Dec. 1963), pp. 599–609.
- [292] R. McMahon, ed. *Avidin-biotin interactions. Methods and applications*. Humana Press, 2008.
- [293] G. T. Hermanson. *Bioconjugate Techniques*. 3rd edition. Elsevier Academic Press, 2013.
- [294] R. Craig, J. C. Cortens, D. Fenyo, and R. C. Beavis. “Using annotated peptide mass spectrum libraries for protein identification”. In: *J. Proteome Res.* 5.8 (Aug. 2006), pp. 1843–1849.
- [295] L. Kuhlmann. “Identification and validation of novel protein biomarkers in pancreatic ductal adenocarcinoma with the ability to distinguish molecular subtypes”. PhD thesis. University of Heidelberg, Germany, 2014.
- [296] A. S. Benk. “Proteomics-based discovery of novel vascular accessible markers in kidney cancer”. PhD thesis. University of Heidelberg, Germany, 2014.

- [297] S. Murugesan, J. Xie, and R. J. Linhardt. “Immobilization of heparin: approaches and applications”. In: *Curr. Top. Med. Chem.* 8.2 (2008), pp. 80–100.
- [298] B. E. Rothenberg, B. K. Hayes, D. Toomre, A. E. Manzi, and A. Varki. “Biotinylated diaminopyridine: an approach to tagging oligosaccharides and exploring their biology”. In: *Proc. Natl. Acad. Sci. USA.* 90.24 (Dec. 1993), pp. 11939–11943.
- [299] C. Leteux, R. A. Childs, W. Chai, M. S. Stoll, H. Kogelberg, and T. Feizi. “Biotinyl-l-3-(2-naphthyl)-alanine hydrazide derivatives of N-glycans: versatile solid-phase probes for carbohydrate-recognition studies”. In: *Glycobiology* 8.3 (Mar. 1998), pp. 227–236.
- [300] N. Finn, M. Quibell, M. Ramjee, and W. Turnell. *Method of producing conjugate vaccines*. US Patent App. 11/879,342. Sept. 2008. URL: <http://www.google.com/patents/US20080213297>.
- [301] B. G. Davis and A. J. Fairbanks. *Carbohydrate Chemistry*. 1st edition. Oxford University Press, 2002.
- [302] A. Dirksen, S. Dirksen, T. M. Hackeng, and P. E. Dawson. “Nucleophilic catalysis of hydrazone formation and transimination: implications for dynamic covalent chemistry”. In: *J. Am. Chem. Soc.* 128.49 (2006), pp. 15602–15603.
- [303] A. Dirksen and P. E. Dawson. “Rapid oxime and hydrazone ligations with aromatic aldehydes for biomolecular labeling”. In: *Bioconjug. Chem.* 19.12 (Dec. 2008), pp. 2543–2548.
- [304] J.-Y. Byeon, F. T. Limpoco, and R. C. Bailey. “Efficient bioconjugation of protein capture agents to biosensor surfaces using aniline-catalyzed hydrazone ligation”. In: *Langmuir* 26.19 (Oct. 2010), pp. 15430–15435.
- [305] S. J. Yang and H. Zhang. “Glycan analysis by reversible reaction to hydrazide beads and mass spectrometry”. In: *Anal. Chem.* 84.5 (Mar. 2012), pp. 2232–2238.
- [306] H. Zhang and S. Yang. *Glycan and glycopeptide capture and release using reversible hydrazone-based method*. US Patent App. 14/124,288. May 2014. URL: <http://www.google.com/patents/US20140135235>.
- [307] R. R. Vivès, E. Crublet, J. P. Andrieu, J. Gagnon, P. Rousselle, and H. Lortat-Jacob. “A novel strategy for defining critical amino acid residues involved in protein/glycosaminoglycan interactions”. In: *J. Biol. Chem.* 279.52 (Dec. 2004), pp. 54327–54333.
- [308] M. J. Wissink, R. Beernink, J. S. Pieper, A. A. Poot, G. H. M. Engbers, T. Beugeling, W. G. van Aken, and J. Feijen. “Immobilization of heparin to EDC/NHS-crosslinked collagen. Characterization and in vitro evaluation”. In: *Biomaterials* 22.2 (Jan. 2001), pp. 151–163.

- [309] X. H. Wang, D. P. Li, W. J. Wang, Q. L. Feng, F. Z. Cui, Y. X. Xu, and X. H. Song. "Covalent immobilization of chitosan and heparin on PLGA surface". In: *Int. J. Biol. Macromol.* 33.1-3 (Nov. 2003), pp. 95–100.
- [310] H. C. Kolb, M. G. Finn, and K. B. Sharpless. "Click Chemistry: Diverse Chemical Function from a Few Good Reactions". In: *Angew. Chem. Int. Ed.* 40.11 (June 2001), pp. 2004–2021.
- [311] V. V. Rostovtsev, L. G. Green, V. V. Fokin, and K. B. Sharpless. "A stepwise huisgen cycloaddition process: copper(I)-catalyzed regioselective "ligation" of azides and terminal alkynes". In: *Angew. Chem. Int. Ed.* 41.14 (July 2002), pp. 2596–2599.
- [312] C. W. Tornøe, C. Christensen, and M. Meldal. "Peptidotriazoles on solid phase: [1,2,3]-triazoles by regiospecific copper(i)-catalyzed 1,3-dipolar cycloadditions of terminal alkynes to azides". In: *J. Org. Chem.* 67.9 (May 2002), pp. 3057–3064.
- [313] M. Meldal and C. W. Tornøe. "Cu-catalyzed azide-alkyne cycloaddition". In: *Chem. Rev.* 108.8 (Aug. 2008), pp. 2952–3015.
- [314] D. C. Dieterich, J. J. Lee, A. J. Link, J. Graumann, D. A. Tirrell, and E. M. Schuman. "Labeling, detection and identification of newly synthesized proteomes with bioorthogonal non-canonical amino-acid tagging". In: *Nat. Protoc.* 2.3 (2007), pp. 532–540.
- [315] J. T. Ngo, J. A. Champion, A. Mahdavi, I. C. Tanrikulu, K. E. Beatty, R. E. Connor, T. H. Yoo, D. C. Dieterich, E. M. Schuman, and D. A. Tirrell. "Cell-selective metabolic labeling of proteins". In: *Nat. Chem. Biol.* 5.10 (Oct. 2009), pp. 715–717.
- [316] C. Y. Jao, M. Roth, R. Welti, and A. Salic. "Metabolic labeling and direct imaging of choline phospholipids in vivo". In: *Proc. Natl. Acad. Sci. U.S.A.* 106.36 (Sept. 2009), pp. 15332–15337.
- [317] A. Salic and T. J. Mitchison. "A chemical method for fast and sensitive detection of DNA synthesis in vivo". In: *Proc. Natl. Acad. Sci. U.S.A.* 105.7 (Feb. 2008), pp. 2415–2420.
- [318] S. T. Laughlin and C. R. Bertozzi. "Imaging the glycome". In: *Proc. Natl. Acad. Sci. U.S.A.* 106.1 (Jan. 2009), pp. 12–17.
- [319] S. R. Hanson, T.-L. Hsu, E. Weerapana, K. Kishikawa, G. M. Simon, B. F. Cravatt, and C.-H. Wong. "Tailored glycoproteomics and glycan site mapping using saccharide-selective bioorthogonal probes". In: *J. Am. Chem. Soc.* 129.23 (June 2007), pp. 7266–7267.
- [320] Y.-Y. Yang, J. M. Ascano, and H. C. Hang. "Bioorthogonal chemical reporters for monitoring protein acetylation". In: *J. Am. Chem. Soc.* 132.11 (Mar. 2010), pp. 3640–3641.

- [321] V. E. Lee, J. M. Schulman, E. I. Stiefel, and C. C. Lee. “Reversible precipitation of bovine serum albumin by metal ions and synthesis, structure and reactivity of new tetrathiometalate chelating agents”. In: *J. Inorg. Biochem.* 101.11-12 (Nov. 2007), pp. 1707–1718.
- [322] R. Agarwal and M. N. Gupta. “Copper affinity precipitation as an initial step in protein purification”. In: *Biotechnol. Tech.* 8.9 (Sept. 1994), pp. 655–658.
- [323] C. Besanceney-Webler, H. Jiang, T. Zheng, L. Feng, D. Soriano del Amo, W. Wang, L. M. Klivansky, F. L. Marlow, Y. Liu, and P. Wu. “Increasing the efficacy of bioorthogonal click reactions for bioconjugation: a comparative study”. In: *Angew. Chem. Int. Ed.* 50.35 (Aug. 2011), pp. 8051–8056.
- [324] D. Soriano Del Amo, W. Wang, H. Jiang, C. Besanceney, A. C. Yan, M. Levy, Y. Liu, F. L. Marlow, and P. Wu. “Biocompatible copper(I) catalysts for in vivo imaging of glycans”. In: *J. Am. Chem. Soc.* 132.47 (Dec. 2010), pp. 16893–16899.
- [325] N. J. Agard, J. A. Prescher, and C. R. Bertozzi. “A strain-promoted [3 + 2] azide-alkyne cycloaddition for covalent modification of biomolecules in living systems”. In: *J. Am. Chem. Soc.* 126.46 (Nov. 2004), pp. 15046–15047.
- [326] J. C. Jewett, E. M. Sletten, and C. R. Bertozzi. “Rapid Cu-free click chemistry with readily synthesized biarylazacyclooctynones”. In: *J. Am. Chem. Soc.* 132.11 (Mar. 2010), pp. 3688–3690.
- [327] J. M. Baskin, J. A. Prescher, S. T. Laughlin, N. J. Agard, P. V. Chang, I. A. Miller, A. Lo, J. A. Codelli, and C. R. Bertozzi. “Copper-free click chemistry for dynamic in vivo imaging”. In: *Proc. Natl. Acad. Sci. U.S.A.* 104.43 (Oct. 2007), pp. 16793–16797.
- [328] P. V. Chang, J. A. Prescher, E. M. Sletten, J. M. Baskin, I. A. Miller, N. J. Agard, A. Lo, and C. R. Bertozzi. “Copper-free click chemistry in living animals”. In: *Proc. Natl. Acad. Sci. U.S.A.* 107.5 (Feb. 2010), pp. 1821–1826.
- [329] J. J. Lemasters, J. DiGuseppi, A.-L. Nieminen, and B. Herman. “Blebbing, free Ca²⁺ and mitochondrial membrane potential preceding cell death in hepatocytes”. In: *Nature* 325.6099 (1987), pp. 78–81.
- [330] G. Zahrebelski, A.-L. Nieminen, K. al-Ghoul, T. Qian, B. Herman, and J. J. Lemasters. “Progression of subcellular changes during chemical hypoxia to cultured rat hepatocytes: a laser scanning confocal microscopic study”. In: *Hepatology* 21.5 (May 1995), pp. 1361–1372.
- [331] J. Chen, X. Liu, L. J. Mandel, and R. G. Schnellmann. “Progressive disruption of the plasma membrane during renal proximal tubule cellular injury”. In: *Toxicol. Appl. Pharmacol.* 171.1 (Feb. 2001), pp. 1–11.

- [332] M.-A. de Ménorval, L. M. Mir, M. L. Fernández, and R. Reigada. “Effects of dimethyl sulfoxide in cholesterol-containing lipid membranes: a comparative study of experiments in silico and with cells”. In: *PLoS ONE* 7.7 (2012), e41733.
- [333] H. Wang, C.-Y. Zhong, J.-F. Wu, Y.-B. Huang, and C.-B. Liu. “Enhancement of TAT cell membrane penetration efficiency by dimethyl sulphoxide”. In: *J. Control. Release* 143.1 (Apr. 2010), pp. 64–70.
- [334] G. Da Violante, N. Zerrouk, I. Richard, G. Provot, J. C. Chaumeil, and P. Arnaud. “Evaluation of the cytotoxicity effect of dimethyl sulfoxide (DMSO) on Caco2/TC7 colon tumor cell cultures”. In: *Biol. Pharm. Bull.* 25.12 (Dec. 2002), pp. 1600–1603.
- [335] I. V. Shilov, S. L. Seymour, A. A. Patel, A. Loboda, W. H. Tang, S. P. Keating, C. L. Hunter, L. M. Nuwaysir, and D. A. Schaeffer. “The Paragon Algorithm, a next generation search engine that uses sequence temperature values and feature probabilities to identify peptides from tandem mass spectra”. In: *Mol. Cell. Proteomics* 6.9 (Sept. 2007), pp. 1638–1655.
- [336] J. R. Knowles. “The mechanism of biotin-dependent enzymes”. In: *Annu. Rev. Biochem.* 58 (1989), pp. 195–221.
- [337] J. Zempleni, S. S. Wijeratne, and Y. I. Hassan. “Biotin”. In: *Biofactors* 35.1 (2009), pp. 36–46.
- [338] S. H. Lecker, A. L. Goldberg, and W. E. Mitch. “Protein degradation by the ubiquitin-proteasome pathway in normal and disease states”. In: *J. Am. Soc. Nephrol.* 17.7 (July 2006), pp. 1807–1819.
- [339] D. Voges, P. Zwickl, and W. Baumeister. “The 26S proteasome: a molecular machine designed for controlled proteolysis”. In: *Annu. Rev. Biochem.* 68 (1999), pp. 1015–1068.
- [340] W. Baumeister, J. Walz, F. Zühl, and E. Seemüller. “The proteasome: paradigm of a self-compartmentalizing protease”. In: *Cell* 92.3 (Feb. 1998), pp. 367–380.
- [341] J. Zempleni. “Uptake, localization, and noncarboxylase roles of biotin”. In: *Annu. Rev. Nutr.* 25 (2005), pp. 175–196.
- [342] N. D. Cohen and M. Thomas. “Biotin transport into fully differentiated 3T3-L1 cells”. In: *Biochem. Biophys. Res. Commun.* 108.4 (Oct. 1982), pp. 1508–1516.
- [343] R. J. McMahon. “Biotin in metabolism and molecular biology”. In: *Annu. Rev. Nutr.* 22 (2002), pp. 221–239.
- [344] L. C. Andersen, J. S. Beaty, J. W. Nettles, C. E. Seyfried, G. T. Nepom, and B. S. Nepom. “Allelic polymorphism in transcriptional regulatory regions of HLA-DQB genes”. In: *J. Exp. Med.* 173.1 (Jan. 1991), pp. 181–192.

- [345] C. Wu and S. Wang. “A pH-sensitive heparin-binding sequence from Baculovirus gp64 protein is important for binding to mammalian cells but not to Sf9 insect cells”. In: *J. Virol.* 86.1 (Jan. 2012), pp. 484–491.
- [346] M. P. Molloy. “Isolation of bacterial cell membranes proteins using carbonate extraction”. In: *Methods Mol. Biol.* 424 (2008), pp. 397–401.
- [347] B. E. Cham and B. R. Knowles. “A solvent system for delipidation of plasma or serum without protein precipitation”. In: *J. Lipid Res.* 17.2 (Mar. 1976), pp. 176–181.
- [348] K. R. Rebecchi, E. P. Go, L. Xu, C. L. Woodin, M. Mure, and H. Desaire. “A general protease digestion procedure for optimal protein sequence coverage and post-translational modifications analysis of recombinant glycoproteins: application to the characterization of human lysyl oxidase-like 2 glycosylation”. In: *Anal. Chem.* 83.22 (Nov. 2011), pp. 8484–8491.
- [349] N. A. Kulak, G. Pichler, I. Paron, N. Nagaraj, and M. Mann. “Minimal, encapsulated proteomic-sample processing applied to copy-number estimation in eukaryotic cells”. In: *Nat. Methods* 11.3 (Mar. 2014), pp. 319–324.
- [350] O. S. Gabrielsen, E. Hornes, L. Korsnes, A. Ruet, and T. B. Øyen. “Magnetic DNA affinity purification of yeast transcription factor tau—a new purification principle for the ultrarapid isolation of near homogeneous factor”. In: *Nucleic Acids Res.* 17.15 (Aug. 1989), pp. 6253–6267.
- [351] A. A. Farooqui. “Purification of enzymes by heparin-sepharose affinity chromatography”. In: *J. Chromatogr.* 184.3 (Aug. 1980), pp. 335–345.
- [352] G. Mitra, E. Hall, and I. Mitra. “Application of immobilized heparins for isolation of human antithrombin III”. In: *Biotechnol. Bioeng.* 28.2 (Feb. 1986), pp. 217–222.
- [353] A. A. Farooqui, H. C. Yang, and L. A. Horrocks. “Purification of lipases, phospholipases and kinases by heparin-Sepharose chromatography”. In: *J. Chromatogr. A* 673.2 (July 1994), pp. 149–158.
- [354] D. Wiese and K. Schmitz. “Expression of recombinant human interleukin-8 and its purification using a single buffer system”. In: *J. Immunol. Methods* 364.1-2 (Feb. 2011), pp. 77–82.
- [355] N. R. Steffen, S. D. Murphy, L. Toller, G. W. Hatfield, and R. H. Lathrop. “DNA sequence and structure: direct and indirect recognition in protein-DNA binding”. In: *Bioinformatics* 18 Suppl 1 (2002), pp. 22–30.
- [356] A. Höglund and O. Kohlbacher. “From sequence to structure and back again: approaches for predicting protein-DNA binding”. In: *Proteome Sci.* 2.1 (June 2004), p. 3.
- [357] S. Ahmad and A. Sarai. “Moment-based prediction of DNA-binding proteins”. In: *J. Mol. Biol.* 341.1 (July 2004), pp. 65–71.

- [358] H. Gadgil, S. A. Oak, and H. W. Jarrett. “Affinity purification of DNA-binding proteins”. In: *J. Biochem. Biophys. Methods* 49.1-3 (Oct. 2001), pp. 607–624.
- [359] H. Gadgil, L. A. Jurado, and H. W. Jarrett. “DNA affinity chromatography of transcription factors”. In: *Anal. Biochem.* 290.2 (Mar. 2001), pp. 147–178.
- [360] E. Genersch, C. Eckerskorn, F. Lottspeich, C. Herzog, K. Kühn, and E. Pöschl. “Purification of the sequence-specific transcription factor CTCBF, involved in the control of human collagen IV genes: subunits with homology to Ku antigen”. In: *EMBO J.* 14.4 (Feb. 1995), pp. 791–800.
- [361] J. Folkman, K. Watson, D. Ingber, and D. Hanahan. “Induction of angiogenesis during the transition from hyperplasia to neoplasia”. In: *Nature* 339.6219 (May 1989), pp. 58–61.
- [362] B. A. W. Hoeben, J. D. M. Molkenboer-Kuenen, W. J. G. Oyen, W. J. M. Peeters, J. H. A. M. Kaanders, J. Bussink, and O. C. Boerman. “Radiolabeled cetuximab: dose optimization for epidermal growth factor receptor imaging in a head-and-neck squamous cell carcinoma model”. In: *Int. J. Cancer* 129.4 (Aug. 2011), pp. 870–878.
- [363] R. A. Herbertson, N. C. Tebbutt, F.-T. Lee, D. J. MacFarlane, B. Chappell, N. Micallef, S.-T. Lee, T. Saunderson, W. Hopkins, F. E. Smyth, D. K. Wyld, J. Bellen, D. S. Sonnichsen, M. W. Brechbiel, C. Murone, and A. M. Scott. “Phase I biodistribution and pharmacokinetic study of Lewis Y-targeting immunoconjugate CMD-193 in patients with advanced epithelial cancers”. In: *Clin. Cancer Res.* 15.21 (Nov. 2009), pp. 6709–6715.
- [364] S. Welt, C. R. Divgi, A. M. Scott, P. Garin-Chesa, R. D. Finn, M. Graham, E. A. Carswell, A. Cohen, S. M. Larson, L. J. Old, and R. W. J. “Antibody targeting in metastatic colon cancer: a phase I study of monoclonal antibody F19 against a cell-surface protein of reactive tumor stromal fibroblasts”. In: *J. Clin. Oncol.* 12.6 (June 1994), pp. 1193–1203.
- [365] J. Folkman. “Clinical applications of research on angiogenesis”. In: *N. Engl. J. Med.* 333 (Dec. 1995), pp. 1757–1762.
- [366] M. S. Dennis, H. Jin, D. Dugger, R. Yang, L. McFarland, A. Ogasawara, S. Williams, M. J. Cole, S. Ross, and R. Schwall. “Imaging tumors with an albumin-binding Fab, a novel tumor-targeting agent”. In: *Cancer Res.* 67.1 (Jan. 2007), pp. 254–261.
- [367] K. L. Gutbrodt, C. Schliemann, L. Giovannoni, K. Frey, T. Pabst, W. Klappper, W. E. Berdel, and D. Neri. “Antibody-based delivery of interleukin-2 to neovasculature has potent activity against acute myeloid leukemia”. In: *Sci. Transl. Med.* 5.201 (Sept. 2013), 201ra118.
- [368] M. Rusckowski, M. Fogarasi, B. Fritz, and D. J. Hnatowich. “Effect of endogenous biotin on the applications of streptavidin and biotin in mice”. In: *Nucl. Med. Biol.* 24.3 (Apr. 1997), pp. 263–268.

- [369] L. Tong. “Structure and function of biotin-dependent carboxylases”. In: *Cell. Mol. Life Sci.* 70.5 (Mar. 2013), pp. 863–891.
- [370] R. Salto, M. D. Girón, M. del Mar Sola, and A. M. Vargas. “Evolution of pyruvate carboxylase and other biotin containing enzymes in developing rat liver and kidney”. In: *Mol. Cell. Biochem.* 200.1-2 (Nov. 1999), pp. 111–117.
- [371] T. Kislinger, B. Cox, A. Kannan, C. Chung, P. Hu, A. Ignatchenko, M. S. Scott, A. O. Gramolini, Q. Morris, M. T. Hallett, J. Rossant, T. R. Hughes, B. Frey, and A. Emili. “Global survey of organ and organelle protein expression in mouse: combined proteomic and transcriptomic profiling”. In: *Cell* 125.1 (Apr. 2006), pp. 173–186.
- [372] E. L. Huttlin, M. P. Jedrychowski, J. E. Elias, T. Goswami, R. Rad, S. A. Beausoleil, J. Villén, W. Haas, M. E. Sowa, and S. P. Gygi. “A tissue-specific atlas of mouse protein phosphorylation and expression”. In: *Cell* 143.7 (Dec. 2010), pp. 1174–1189.
- [373] L. F. Waanders, K. Chwalek, M. Monetti, C. Kumar, E. Lammert, and M. Mann. “Quantitative proteomic analysis of single pancreatic islets”. In: *Proc. Natl. Acad. Sci. U.S.A.* 106.45 (Nov. 2009), pp. 18902–18907.
- [374] J. X. Zhang, I. Braakman, K. E. S. Matlack, and A. Helenius. “Quality control in the secretory pathway: the role of calreticulin, calnexin and BiP in the retention of glycoproteins with C-terminal truncations”. In: *Mol. Biol. Cell* 8.10 (Oct. 1997), pp. 1943–1954.
- [375] M. E. Feder and G. E. Hofmann. “Heat-shock proteins, molecular chaperones, and the stress response: evolutionary and ecological physiology”. In: *Annu. Rev. Physiol.* 61 (1999), pp. 243–282.
- [376] H. Ginisty, H. Sicard, B. Roger, and P. Bouvet. “Structure and functions of nucleolin”. In: *J. Cell. Sci.* 112 (Mar. 1999), pp. 761–772.
- [377] M. M. Tajrishi, R. Tuteja, and N. Tuteja. “Nucleolin: The most abundant multifunctional phosphoprotein of nucleolus”. In: *Commun. Integr. Biol.* 4.3 (May 2011), pp. 267–275.
- [378] D. J. Taylor, B. Devkota, A. D. Huang, M. Topf, E. Narayanan, A. Sali, S. C. Harvey, and J. Frank. “Comprehensive molecular structure of the eukaryotic ribosome”. In: *Structure* 17.12 (Dec. 2009), pp. 1591–1604.
- [379] B. M. Gumbiner. “Regulation of cadherin-mediated adhesion in morphogenesis”. In: *Nat. Rev. Mol. Cell Biol.* 6.8 (Aug. 2005), pp. 622–634.
- [380] J. K. Mouw, G. Ou, and V. M. Weaver. “Extracellular matrix assembly: a multiscale deconstruction”. In: *Nat. Rev. Mol. Cell Biol.* 15.12 (Dec. 2014), pp. 771–785.

- [381] A. Calvo, R. Catena, M. S. Noble, D. Carbott, I. Gil-Bazo, O. Gonzalez-Moreno, J.-I. Huh, R. Sharp, T.-H. Qiu, M. R. Anver, G. Merlino, R. B. Dickson, M. D. Johnson, and J. E. Green. “Identification of VEGF-regulated genes associated with increased lung metastatic potential: functional involvement of tenascin-C in tumor growth and lung metastasis”. In: *Oncogene* 27.40 (Sept. 2008), pp. 5373–5384.
- [382] A. Karihaloo, S. Kale, N. D. Rosenblum, and L. G. Cantley. “Hepatocyte growth factor-mediated renal epithelial branching morphogenesis is regulated by glypican-4 expression”. In: *Mol. Cell. Biol.* 24.19 (Oct. 2004), pp. 8745–8752.
- [383] Y. Takai, K. Irie, K. Shimizu, T. Sakisaka, and W. Ikeda. “Nectins and nectin-like molecules: roles in cell adhesion, migration, and polarization”. In: *Cancer Sci.* 94.8 (Aug. 2003), pp. 655–667.
- [384] I. D. Campbell and M. J. Humphries. “Integrin structure, activation, and interactions”. In: *Cold Spring Harb. Perspect. Biol.* 3.3 (Mar. 2011).
- [385] H.-X. Yan, W. Yang, R. Zhang, L. Chen, L. Tang, B. Zhai, S.-Q. Liu, H.-F. Cao, X.-B. Man, H.-P. Wu, M.-C. Wu, and H.-Y. Wang. “Protein-tyrosine phosphatase PCP-2 inhibits beta-catenin signaling and increases E-cadherin-dependent cell adhesion”. In: *J. Biol. Chem.* 281.22 (June 2006), pp. 15423–15433.
- [386] M. Krüger, M. Moser, S. Ussar, I. Thievensen, C. A. Luber, F. Forner, S. Schmidt, S. Zanivan, R. Fässler, and M. Mann. “SILAC mouse for quantitative proteomics uncovers kindlin-3 as an essential factor for red blood cell function”. In: *Cell* 134.2 (July 2008), pp. 353–364.
- [387] J. Wang, F. Gao, F. Mo, X. Hong, H. Wang, S. Zheng, and B. Lin. “Identification of CHI3L1 and MASP2 as a biomarker pair for liver cancer through integrative secretome and transcriptome analysis”. In: *Proteom. Clin. Appl.* 3 (Oct. 2009), pp. 541–551.
- [388] R. Shi, C. Kumar, A. Zougman, Y. Zhang, A. Podtelejnikov, J. Cox, J. R. Wiśniewski, and M. Mann. “Analysis of the mouse liver proteome using advanced mass spectrometry”. In: *J. Proteome Res.* 6.8 (Aug. 2007), pp. 2963–2972.
- [389] X. Li, D. Han, R. Kin Ting Kam, X. Guo, M. Chen, Y. Yang, H. Zhao, and Y. Chen. “Developmental expression of sideroflexin family genes in *Xenopus* embryos”. In: *Dev. Dyn.* 239.10 (Oct. 2010), pp. 2742–2747.
- [390] A. Azzi, M. Glerum, R. Koller, W. Mertens, and S. Spycher. “The mitochondrial tricarboxylate carrier”. In: *J. Bioenerg. Biomembr.* 25.5 (Oct. 1993), pp. 515–524.

- [391] M. D. Fleming, D. R. Campagna, J. N. Haslett, C. C. Trenor III, and N. C. Andrews. “A mutation in a mitochondrial transmembrane protein is responsible for the pleiotropic hematological and skeletal phenotype of flexed-tail (f/f) mice”. In: *Genes Dev.* 15.6 (Mar. 2001), pp. 652–657.
- [392] P. J. Lockhart, B. Holtom, S. Lincoln, J. Hussey, A. Zimprich, T. Gasser, Z. K. Wszolek, J. Hardy, and M. J. Farrer. “The human sideroflexin 5 SFXN5 gene: sequence, expression analysis and exclusion as a candidate for PARK3”. In: *Gene* 285.1-2 (Feb. 2002), pp. 229–237.
- [393] H. Zaid, S. Abu-Hamad, A. Israelson, I. Nathan, and V. Shoshan-Barmatz. “The voltage-dependent anion channel-1 modulates apoptotic cell death”. In: *Cell Death Differ.* 12.7 (July 2005), pp. 751–760.
- [394] V. De Pinto, A. Messina, D. J. R. Lane, and A. Lawen. “Voltage-dependent anion-selective channel (VDAC) in the plasma membrane”. In: *FEBS Lett.* 584.9 (May 2010), pp. 1793–1799.
- [395] S. Mechanic, K. Raynor, J. E. Hill, and P. Cowin. “Desmocollins form a distinct subset of the cadherin family of cell adhesion molecules”. In: *Proc. Natl. Acad. Sci. U.S.A.* 88.10 (May 1991), pp. 4476–4480.
- [396] V. Maier, C. Jolicoeur, H. Rayburn, N. Takegahara, A. Kumanogoh, H. Kikutani, M. Tessier-Lavigne, W. Wurst, and R. H. Friedel. “Semaphorin 4C and 4G are ligands of Plexin-B2 required in cerebellar development”. In: *Mol. Cell. Neurosci.* 46.2 (Feb. 2011), pp. 419–431.
- [397] N. Echols, P. Harrison, S. Balasubramanian, N. M. Luscombe, P. Bertone, Z. Zhang, and M. Gerstein. “Comprehensive analysis of amino acid and nucleotide composition in eukaryotic genomes, comparing genes and pseudogenes”. In: *Nucleic Acids Res.* 30.11 (June 2002), pp. 2515–2523.
- [398] A. B. Lowe. “Thiol-yne ‘click’/coupling chemistry and recent applications in polymer and materials synthesis and modification”. In: *Polymer* 55 (Aug. 2014), pp. 5517–5549.
- [399] R. Ekkebus, S. I. van Kasteren, Y. Kulathu, A. Scholten, I. Berlin, P. P. Geurink, A. de Jong, S. Goerdal, J. Neefjes, A. J. R. Heck, D. Komander, and H. Ova. “On terminal alkynes that can react with active-site cysteine nucleophiles in proteases”. In: *J. Am. Chem. Soc.* 135.8 (Feb. 2013), pp. 2867–2870.
- [400] R. L. Gundry, M. Y. White, C. I. Murray, L. A. Kane, Q. Fu, B. A. Stanley, and J. E. Van Eyk. “Preparation of proteins and peptides for mass spectrometry analysis in a bottom-up proteomics workflow”. In: *Curr. Protoc. Mol. Biol.* Chapter 10 (Oct. 2009), Unit 10.25.
- [401] E. S. Boja and H. M. Fales. “Overalkylation of a protein digest with iodoacetamide”. In: *Anal. Chem.* 73.15 (Aug. 2001), pp. 3576–3582.

- [402] T. Vaisar. “Thematic review series: proteomics. Proteomic analysis of lipid-protein complexes”. In: *J. Lipid Res.* 50.5 (May 2009), pp. 781–786.
- [403] F. H. Carpenter. “Treatment of trypsin with TPCK”. In: *Method. Enzymol.* 11 (1967), p. 237.
- [404] T. L. Constantopoulos, G. S. Jackson, and C. G. Enke. “Effects of salt concentration on analyte response using electrospray ionization mass spectrometry”. In: *J. Am. Soc. Mass Spectrom.* 10.7 (July 1999), pp. 625–634.
- [405] M. R. Larsen, S. J. Cordwell, and P. Roepstorff. “Graphite powder as an alternative or supplement to reversed-phase material for desalting and concentration of peptide mixtures prior to matrix-assisted laser desorption/ionization-mass spectrometry”. In: *Proteomics* 2.9 (Sept. 2002), pp. 1277–1287.
- [406] M. Palmblad and J. S. Vogel. “Quantitation of binding, recovery and desalting efficiency of peptides and proteins in solid phase extraction micropipette tips”. In: *J. Chromatogr. B* 814.2 (Jan. 2005), pp. 309–313.
- [407] J. Peng and S. P. Gygi. “Proteomics: the move to mixtures”. In: *J. Mass. Spectrom.* 36.10 (Oct. 2001), pp. 1083–1091.
- [408] W.-J. Qian, J. M. Jacobs, T. Liu, D. G. Camp III, and R. D. Smith. “Advances and challenges in liquid chromatography-mass spectrometry-based proteomics profiling for clinical applications”. In: *Mol. Cell. Proteomics* 5.10 (Oct. 2006), pp. 1727–1744.
- [409] Y. Yang, S. Zhang, K. Howe, D. B. Wilson, F. Moser, D. Irwin, and T. W. Thannhauser. “A comparison of nLC-ESI-MS/MS and nLC-MALDI-MS/MS for GeLC-based protein identification and iTRAQ-based shotgun quantitative proteomics”. In: *J. Biomol. Tech.* 18.4 (Sept. 2007), pp. 226–237.
- [410] W. M. Bodnar, R. K. Blackburn, J. M. Krise, and M. A. Moseley. “Exploiting the complementary nature of LC/MALDI/MS/MS and LC/ESI/MS/MS for increased proteome coverage”. In: *J. Am. Soc. Mass Spectrom.* 14.9 (Sept. 2003), pp. 971–979.
- [411] A. Tangen, H. Ehring, J. Flensburg, M. C. P. Conaway, and H. Wadensten. “The LC-MS Workflow: Is MALDI or ESI the Preferable Ionization Method?”. In: *ASMS (Conference Presentation)*. 2004.
- [412] A. I. Nesvizhskii and R. Aebersold. “Interpretation of shotgun proteomic data: the protein inference problem”. In: *Mol. Cell. Proteomics* 4.10 (Oct. 2005), pp. 1419–1440.
- [413] R. Usaite, J. Wohlschlegel, J. D. Venable, S. K. Park, J. Nielsen, L. Olsson, and J. R. Yates III. “Characterization of global yeast quantitative proteome data generated from the wild-type and glucose repression *saccharomyces cerevisiae* strains: the comparison of two quantitative methods”. In: *J. Proteome Res.* 7.1 (Jan. 2008), pp. 266–275.

- [414] S. Jin, D. S. Daly, D. L. Springer, and J. H. Miller. “The effects of shared peptides on protein quantitation in label-free proteomics by LC/MS/MS”. In: *J. Proteome Res.* 7.1 (Jan. 2008), pp. 164–169.
- [415] Y. Zhang, Z. Wen, M. P. Washburn, and L. Florens. “Refinements to label free proteome quantitation: how to deal with peptides shared by multiple proteins”. In: *Anal. Chem.* 82.6 (Mar. 2010), pp. 2272–2281.
- [416] B. Zybaylov, H. Rutschow, G. Friso, A. Rudella, O. Emanuelsson, Q. Sun, and K. J. van Wijk. “Sorting signals, N-terminal modifications and abundance of the chloroplast proteome”. In: *PLoS ONE* 3.4 (2008), e1994.
- [417] D. L. Tabb et al. “Repeatability and reproducibility in proteomic identifications by liquid chromatography-tandem mass spectrometry”. In: *J. Proteome Res.* 9.2 (Feb. 2010), pp. 761–776.
- [418] C. Rösli, J.-N. Rybak, D. Neri, and G. Elia. “Avidin-Biotin Interactions”. In: ed. by R. J. McMahon. Humana Press, 2008. Chap. Quantitative Recovery of Biotinylated Proteins from Streptavidin-Based Affinity Chromatography Resins, pp. 89–100.
- [419] C. Hess, D. Venetz, and D. Neri. “Emerging classes of armed antibody therapeutics against cancer”. In: *Med. Chem. Commun.* 5 (Jan. 2014), pp. 408–431.

A. Supplementary Figures

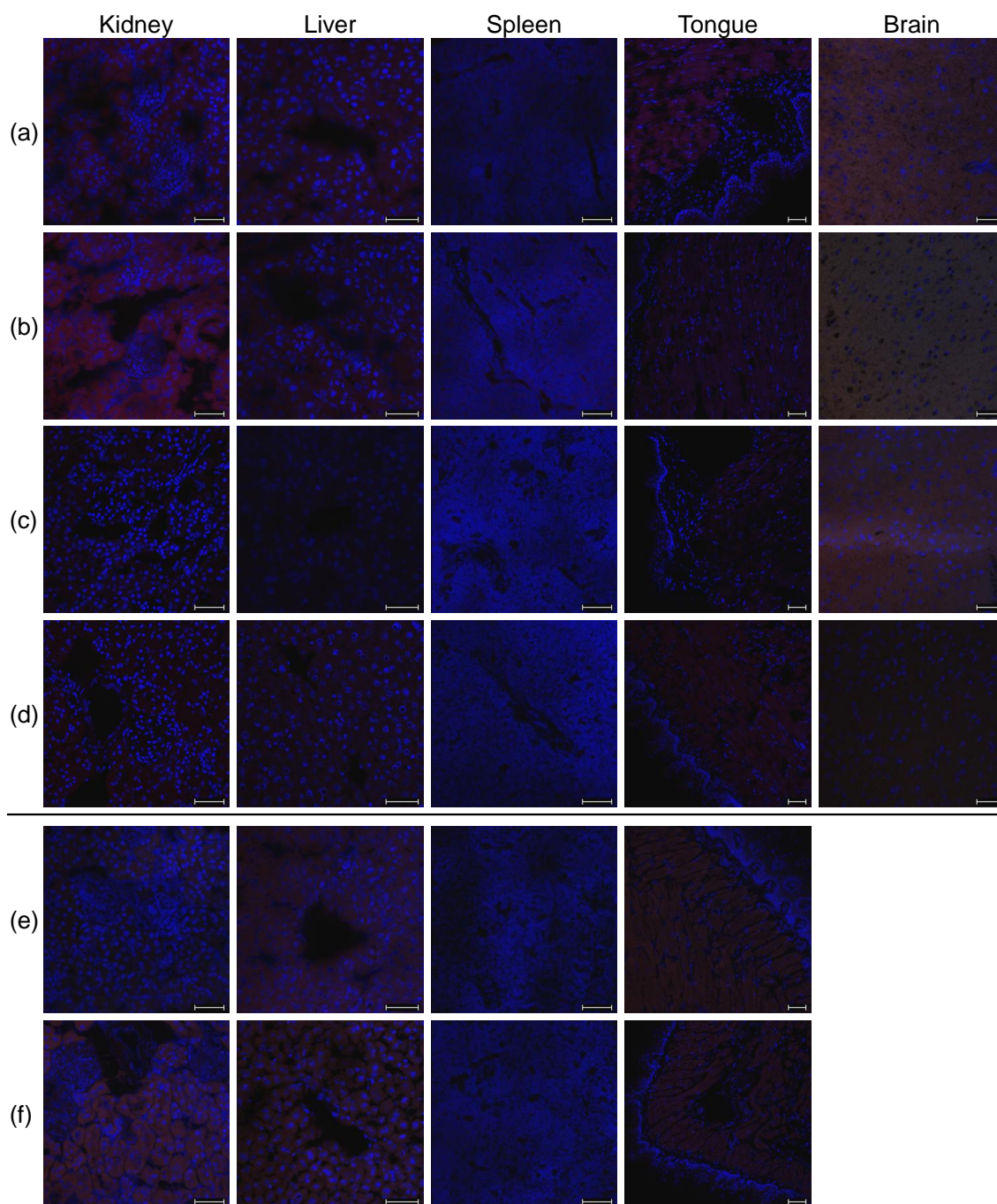


Figure A.1.: CLSM staining controls of mouse tissue perfused with commercial and peptide-based reagents (corresponding to Figure 4.31) as well as with heparin-based reagents (corresponding to Figure 4.32). Scale bars: 50 μm . Sections from mouse tissue perfused with Sulfo-NHS-LC-biotin (a), NHS-PEG₁₂-biotin (b), SMCC-Cys-(L-Asp)₃-biotin (c), SM(PEG)₆-Cys-(L-Asp)₃-biotin (d), Biotin-Heparin-sNHS (e), Biotin-Enoxaparin-sNHS (f).

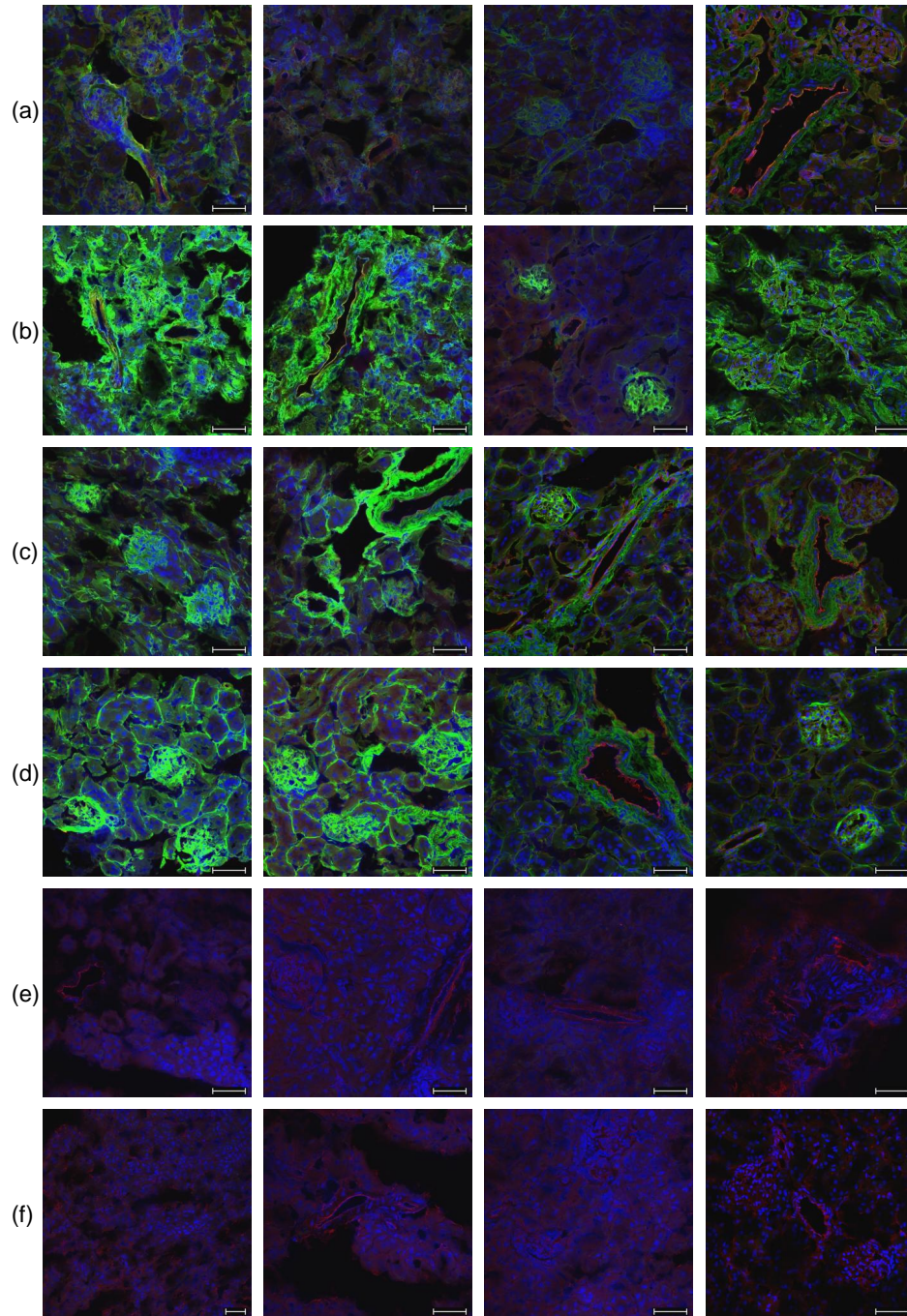


Figure A.2.: CLSM of all mouse kidney samples perfused with commercial and peptide-based reagents used for mass spectrometric *in vivo* validation. DAPI (blue), biotin (green), endothelial marker CD31 (red), scale bars: 50 μm . Perfusion with Sulfo-NHS-LC-biotin (a), NHS-PEG₁₂-biotin (b), SMCC-Cys-(L-Asp)₃-biotin (c), SM(PEG)₆-Cys-(L-Asp)₃-biotin (d), PBS perfused controls (e), non-perfused controls (f). Corresponding staining controls see Suppl. Figure A.4.

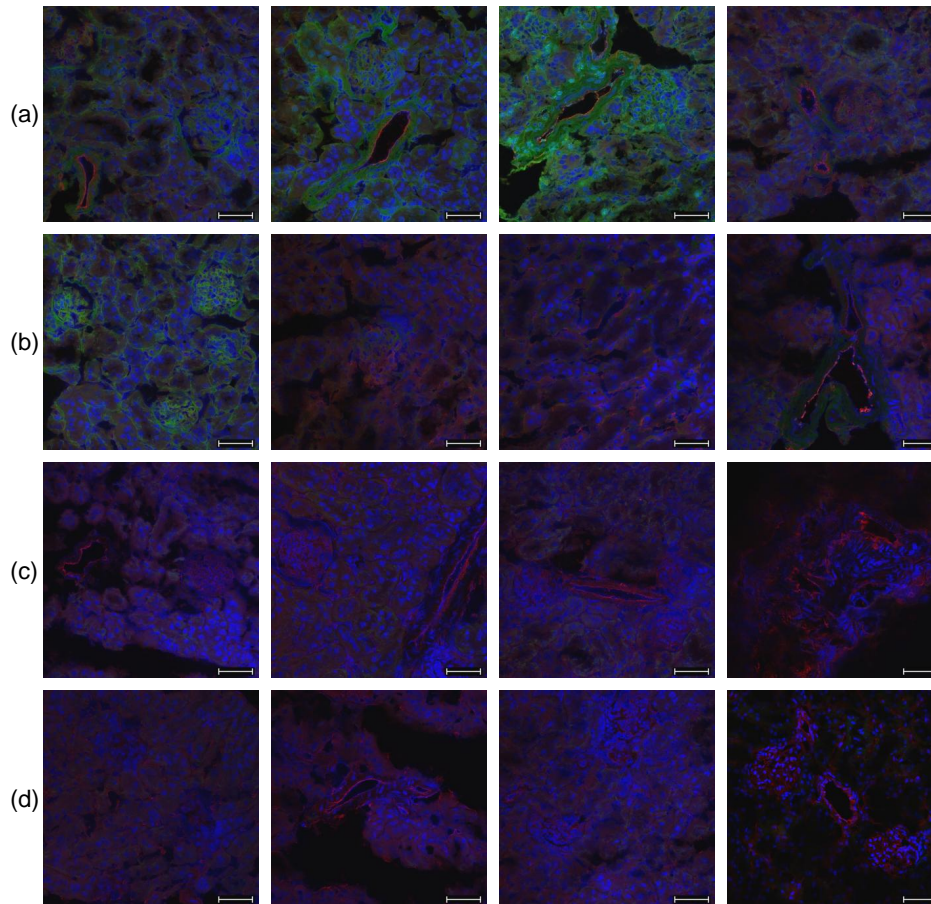


Figure A.3.: CLSM of all mouse kidney samples perfused with heparin-based reagents used for mass spectrometric *in vivo* validation. DAPI (blue), biotin (green), endothelial marker CD31 (red), scale bars: 50 μm . Perfusion with Biotin-Heparin-sNHS (a), Biotin-Enoxaparin-sNHS (b), PBS perfused controls (c), non-perfused controls (d). Corresponding staining controls see Suppl. Figure A.4.

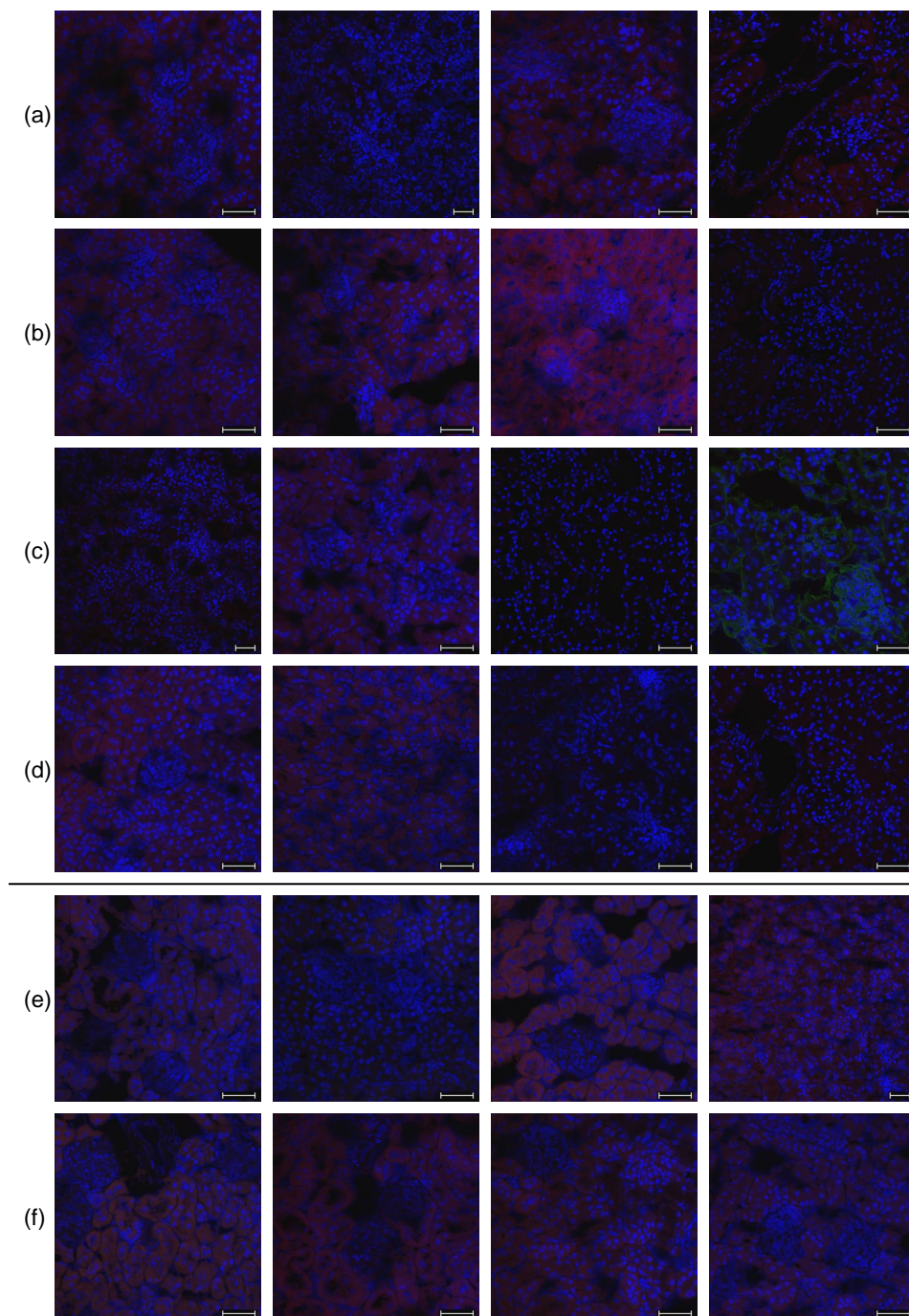


Figure A.4.: CLSM staining controls of kidney tissue perfused with commercial and peptide-based reagents (corresponding to Suppl. Figure A.2) as well as with heparin-based reagents (corresponding to Suppl. Figure A.3). Scale bars: 50 μm . Sections from mouse tissue perfused with Sulfo-NHS-LC-biotin (a), NHS-PEG₁₂-biotin (b), SMCC-Cys-(L-Asp)₃-biotin (c), SM(PEG)₆-Cys-(L-Asp)₃-biotin (d), Biotin-Heparin-sNHS (e), Biotin-Enoxaparin-sNHS (f).

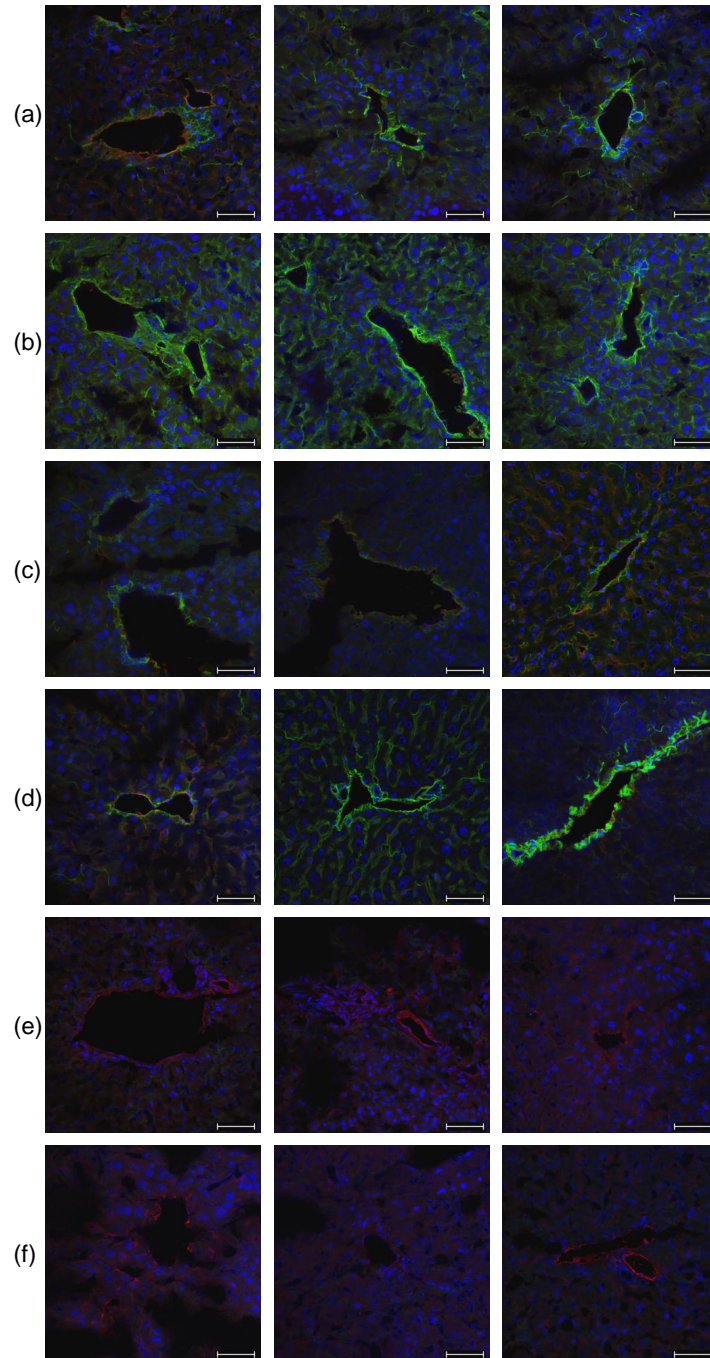


Figure A.5.: CLSM of all mouse liver samples perfused with commercial and peptide-based reagents used for mass spectrometric *in vivo* validation. DAPI (blue), biotin (green), endothelial marker CD31 (red), scale bars: 50 μm . Perfusion with Sulfo-NHS-LC-biotin (a), NHS-PEG₁₂-biotin (b), SMCC-Cys-(L-Asp)₃-biotin (c), SM(PEG)₆-Cys-(L-Asp)₃-biotin (d), PBS perfused controls (e), non-perfused controls (f). Corresponding staining controls see Suppl. Figure A.7.

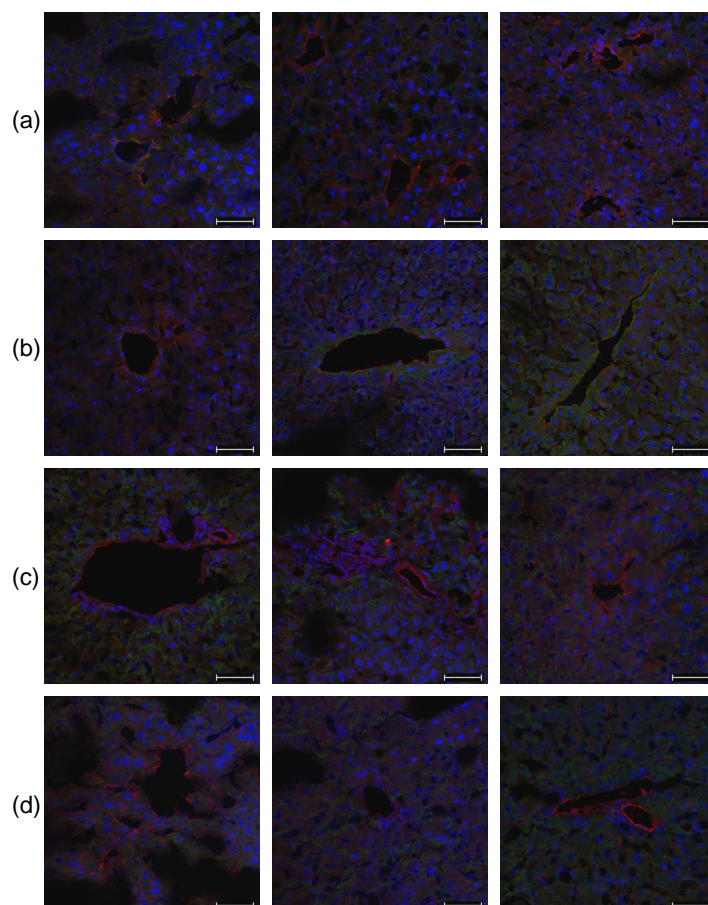


Figure A.6.: CLSM of all mouse liver samples perfused with heparin-based reagents used for mass spectrometric *in vivo* validation. DAPI (blue), biotin (green), endothelial marker CD31 (red), scale bars: 50 μm . Perfusion with Biotin-Heparin-sNHS (a), Biotin-Enoxaparin-sNHS (b), PBS perfused controls (c), non-perfused controls (d). Corresponding staining controls see Suppl. Figure A.7.

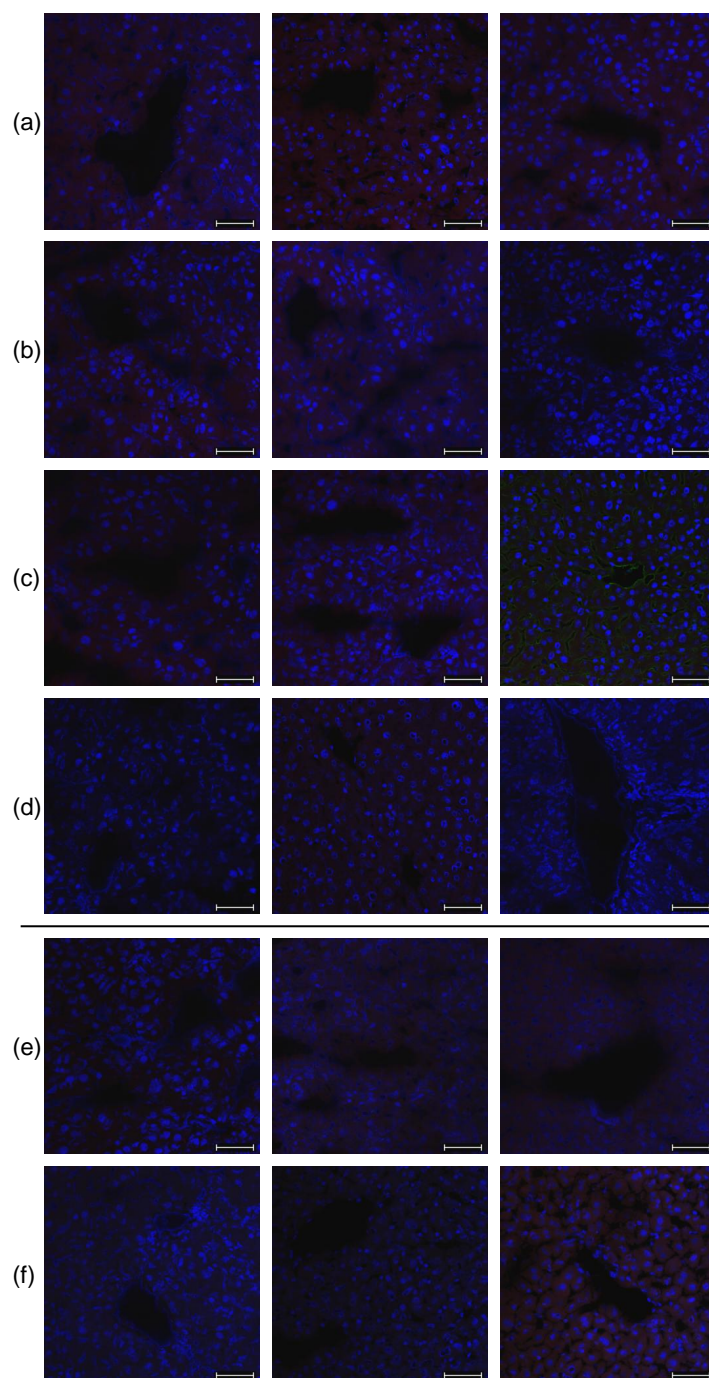


Figure A.7.: CLSM staining controls of liver tissue perfused with commercial and peptide-based reagents (corresponding to Suppl. Figure A.5) as well as with heparin-based reagents (corresponding to Suppl. Figure A.6). Scale bars: 50 μm . Sections from mouse tissue perfused with Sulfo-NHS-LC-biotin (a), NHS-PEG₁₂-biotin (b), SMCC-Cys-(L-Asp)₃-biotin (c), SM(PEG)₆-Cys-(L-Asp)₃-biotin (d), Biotin-Heparin-sNHS (e), Biotin-Enoxaparin-sNHS (f).

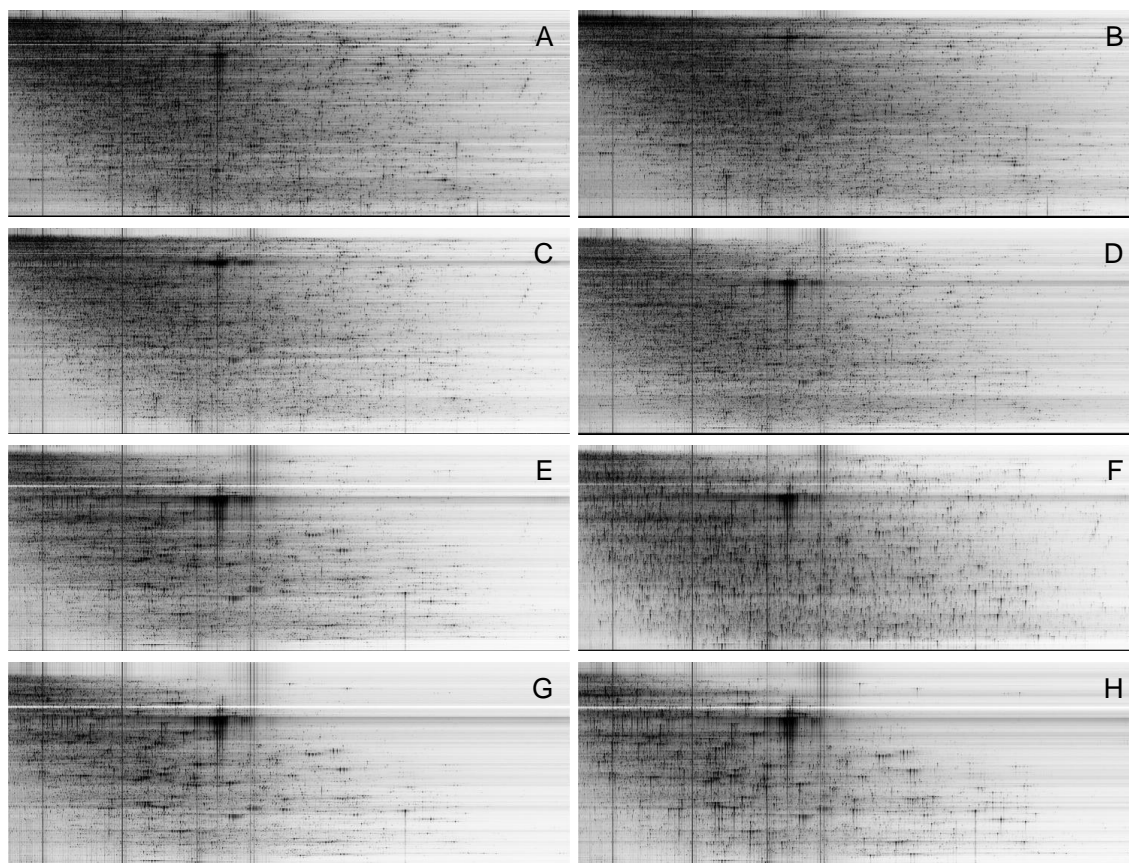


Figure A.8.: 2D peptide maps of *in vivo* biotinylated kidney samples (exemplary maps). Biotinylation with Sulfo-NHS-LC-biotin (A), NHS-PEG₁₂-biotin (B), SMCC-Cys-(L-Asp)₃-biotin (C), SM(PEG)₆-Cys-(L-Asp)₃-biotin (D), Biotin-Heparin-sNHS (E), Biotin-Enoxaparin-sNHS (F), negative controls: PBS (G), non-perfused (H). m/z ratio (x-axis), UPLC fractions (y-axis) and normalized signal intensity as grey scale. Normalization to internal standard at m/z 1411 with standard peak height set to 1000, peak height capping at 500.

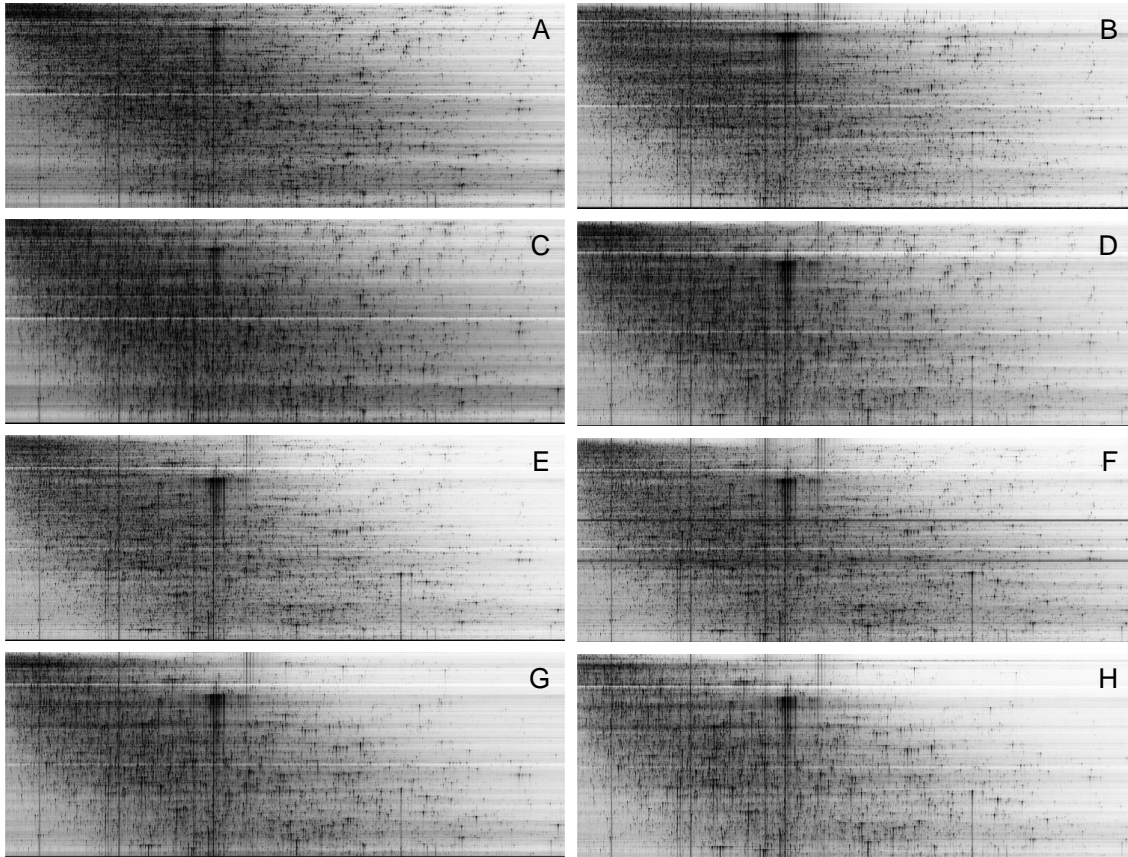


Figure A.9.: 2D peptide maps of *in vivo* biotinylated liver samples (exemplary maps). Biotinylation with Sulfo-NHS-LC-biotin (A), NHS-PEG₁₂-biotin (B), SMCC-Cys-(L-Asp)₃-biotin (C), SM(PEG)₆-Cys-(L-Asp)₃-biotin (D), Biotin-Heparin-sNHS (E), Biotin-Enoxaparin-sNHS (F), negative controls: PBS (G), non-perfused (H). m/z ratio (x-axis), UPLC fractions (y-axis) and normalized signal intensity as grey scale. Normalization to internal standard at m/z 1411 with standard peak height set to 1000, peak height capping at 500.

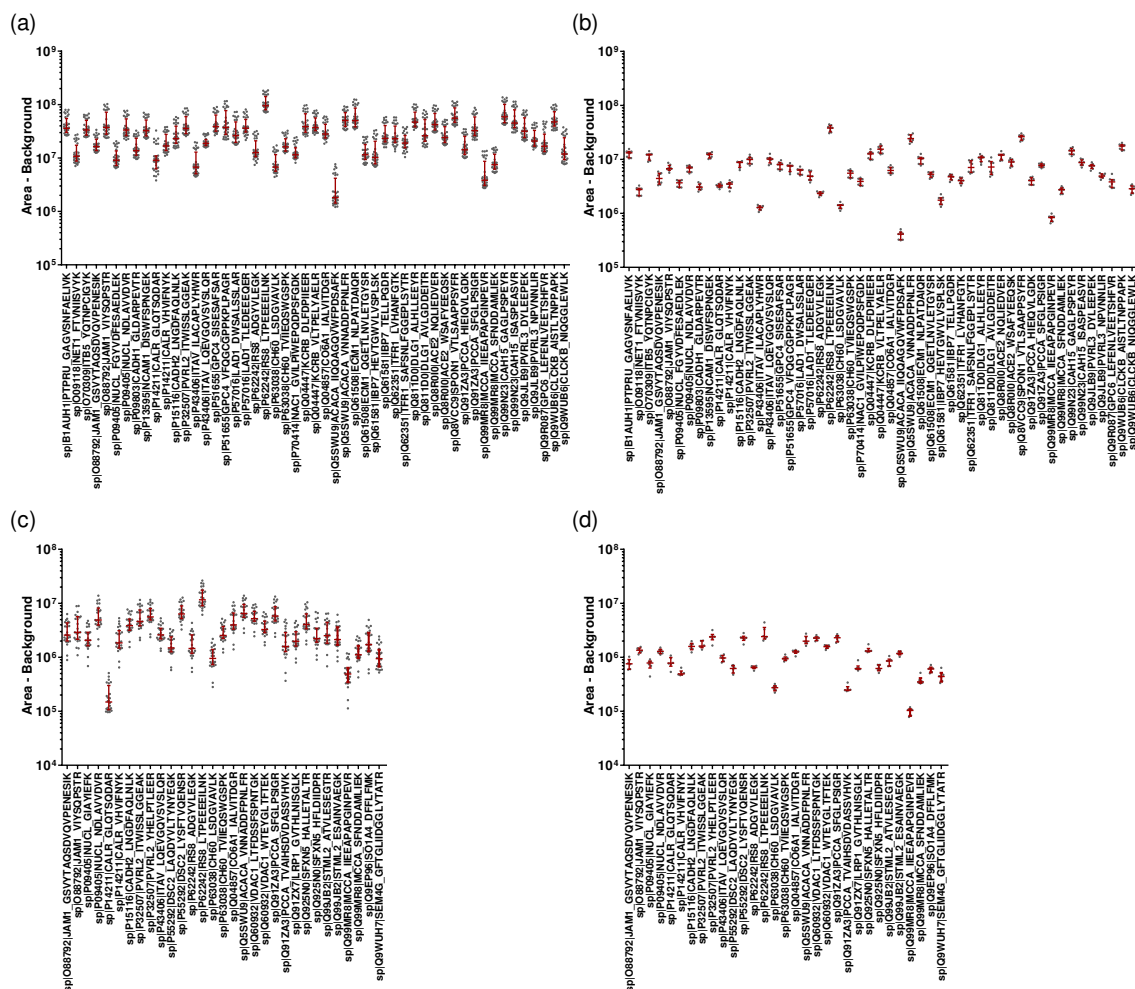


Figure A.10.: Background-reduced peak areas of the heavy standards spiked into all samples, SRM analysis, median with interquartile range. Kidney samples: 32 vascular enriched sample (a) and 8 full proteome samples (b). Liver samples: 24 vascular enriched sample (c) and 6 full proteome samples (d).

B. Supplementary Tables

The Supplementary Tables are enclosed in Microsoft's Office Open XML format (.xlsx) on a CD.

Table B.1.: Transition list SRM analysis, kidney tissue. Proteins, peptide sequences and selected transitions are shown, including information about precursor and fragment m/z, retention time, declustering potential and collision energy.

Table B.2.: Transition list SRM analysis, liver tissue. Proteins, peptide sequences and selected transitions are shown, including information about precursor and fragment m/z, retention time, declustering potential and collision energy.

Table B.3.: *In vitro* validation of peptide-based reagents, analysis of peptide summaries. Numbers of identified proteotypic peptides per reagent group and protein, including localization information (plasma membrane (pm), extracellular and secreted (ex), membrane associated (mem), intracellular (int), no localization information (no)).

Table B.4.: *In vitro* validation of peptide-based reagents, relative quantification of all reagent groups against Sulfo-NHS-LC-biotin, performed with MS_QBAT.

Table B.5.: *In vitro* validation of heparin-based reagents, analysis of peptide summaries. Numbers of identified proteotypic peptides per protein, including localization information (plasma membrane (pm), extracellular and secreted (ex), membrane associated (mem), intracellular (int), no localization information (no)). Duplicate analysis in comparison to Sulfo-NHS-LC-biotin (a), different sample preparation strategies (b), DNA blocking and non-activated Biotin-Heparin (c).

Table B.6.: *In vivo* validation, kidney tissue, analysis of peptide summaries. Numbers of identified proteotypic peptides per reagent group and protein, including localization information (plasma membrane (pm), extracellular and secreted (ex), membrane associated (mem), cytoplasmic site of plasma membrane (ipm), intracellular (int), no localization information (no)).

Table B.7.: *In vivo* validation, kidney tissue, relative quantification of all reagent groups against Sulfo-NHS-LC-biotin, performed with MS_QBAT. Including localization information (plasma membrane (pm), extracellular and secreted (ex), membrane associated (mem), cytoplasmic site of plasma membrane (ipm), intracellular (int), no localization information (no)).

Table B.8.: *In vivo* validation, kidney background proteome analysis, different sample preparation strategies, analysis of peptide summaries. Numbers of identified proteotypic peptides per reagent group and protein, including localization information (plasma membrane (pm), extracellular and secreted (ex), membrane associated (mem), cytoplasmic site of plasma membrane (ipm), intracellular (int), no localization information (no)).

Table B.9.: *In vivo* validation, liver tissue, analysis of peptide summaries. Numbers of identified proteotypic peptides per reagent group and protein, including localization information (plasma membrane (pm), extracellular and secreted (ex), membrane associated (mem), cytoplasmic site of plasma membrane (ipm), intracellular (int), no localization information (no)).

Table B.10.: *In vivo* validation, liver tissue, relative quantification of all reagent groups against Sulfo-NHS-LC-biotin, performed with MS_QBAT. Including localization information (plasma membrane (pm), extracellular and secreted (ex), membrane associated (mem), cytoplasmic site of plasma membrane (ipm), intracellular (int), no localization information (no)).